OPEN ACCESS

University of BRISTOL

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Parsons, Kieran John

Title:
Highly efficient broadband linearisation of RF amplifiers for satellite and cellular applications.

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Highly Efficient Broadband Linearisation of RF Amplifiers for Satellite and Cellular Applications

Kieran John Parsons

March 1996

A thesis submitted to the University of Bristol in accordance with the requirements for the degree of Doctor of Philosophy in the Faculty of Engineering, Department of Electrical and Electronic Engineering.

Abstract

The use of radio links for communications purposes has increased rapidly during recent years. This has led to the requirement for increased spectral efficiency of the communication systems in order to fulfil demand using the scarce resources available. This, in turn, has led to detailed research into the use of linear modulation techniques. In order to preserve the spectral efficiency provided by linear modulation, a linear transmitter is required. In addition, the needs of multi-channel transmitters, such as cellular basestations and satellite transponders could also be met with a single low-distortion linear amplifier.

There is, therefore, a demand for a high quality linear amplifier, with a high degree of power efficiency. Since these properties are usually contradictory, new amplifier topologies have been proposed to fulfil the required specifications.

To this end, this work has centred upon the use of the feedforward and analogue predistortion linearisation techniques upon highly efficient RF amplifiers. It has been shown that neither feedforward linearisation nor predistortion can, in general, provide the required performance individually, but a combined use of the techniques has the potential for attaining excellent performance.

To investigate the applicability of the feedforward linearisation technique, a practical system has been built using highly efficient class C amplifiers. The efficiency of the system has been theoretically and experimentally investigated, and it has been shown that, for the practical system, it is possible to achieve a much higher efficiency than can be attained with conventional amplifiers of similar linearity. However, it has been shown that the insertion loss of the time delays required in the technique degrade this improved efficiency. This problem has been addressed, and a new technique has been derived to increase efficiency whilst sacrificing some linearity. An alternative technique, applicable to highly nonlinear amplifiers, has also been suggested to improve efficiency by recovering some of the power inevitably wasted in the feedforward system.

Analogue predistortion has been used extensively to linearise saturating amplifiers, such as class AB or travelling wave tube amplifiers. However, its use with class C amplifiers does not seem to have been widely studied. A practical predistortion amplifier has been constructed, using a class C power amplifier module. It has been shown that the linearity improvement which can be attained is limited, however its efficiency has not been degraded to the same extent as in the feedforward system. The practical system has been extended to illustrate the combined use of feedforward and predistortion linearisation, and shows that the performance of the composite system is significantly greater than that achieved with either technique in isolation.

To further increase the performance of the combined system, it is necessary to improve the degree of linearity which the predistorter provides. Thus, the use of a novel predistorter element, known as a piecewise linear predistorter, has been investigated by means of computer simulation. It has been shown that this technique offers the potential of much improved linearity compared with standard predistortion architectures.

Author's Declaration

Unless otherwise acknowledged, the content of this thesis is the original and sole work of the author. No portion of the work in this thesis has been submitted by the author in support of an application for any other degree or qualification, at this or any other University or institute of learning. The views expressed in this thesis are those of the author, and not necessarily those of the University of Bristol.

Kieran John Parsons

Copyright

Attention is drawn to the fact that the copyright of this thesis rests with the author. This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author. This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purpose of consultation.

CONTENTS

List of Figures

List of Tables

List of Abbreviations

ac	Alternating current
AF	Audio Frequency
AM	Amplitude Modulation
BJT	Bipolar Junction Transistor
CALLUM	Combined Analogue Locked Loop Universal Modulator
CDMA	Code Division Multiple Access
CW	Carrier Wave
dB	Decibel
dBc	Decibels relative to carrier
dBm	Decibels relative to 1mW
dc	Direct current
$(\pi/4)$-DQPSK	$(\pi/4)$-Differential Quadrature Phase Shift Keying
DS-CDMA	Direct Sequence Code Division Multiple Access
DSP	Digital Signal Processor(ing)
EE&R	Envelope Elimination and Restoration
FET	Field Effect Transistor
FFT	Fast Fourier Transform
GaAs FET	Gallium Arsenide Field Effect Transistor
G-PLP	Gain-Type Piecewise Linear Predistorter
HF	High Frequency
IF	Intermediate Frequency
IMD	Intermodulation Distortion

IMP	Intermodulation Distortion Product
JDC	Japanese Digital Cellular
LINC	LInear amplification using Nonlinear Components
LO	Local Oscillator
MMIC	Monolithic Microwave Integrated Circuit
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
NADC	North American Digital Cellular
p.d.f.	Probability Density Function
PEP	Peak Envelope Power
PLP	Piecewise Linear Predistorter
PM	Phase Modulation
PN	Pseudo–Noise
PWM	Pulse Width Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RFC	Radio Frequency Choke
TC–PLP	Transfer Characteristic–Type Piecewise Linear Predistorter
TETRA	Trans–European Trunked RAdio
TWTA	Travelling Wave Tube Amplifier
UGBW	Unity Gain Bandwidth
UHF	Ultra High Frequency
VCO	Voltage Controlled Oscillator
VHF	Very High Frequency

PUBLICATIONS

[1] P.B. KENINGTON, K.J. PARSONS, AND D.W. BENNETT. Broadband linearisation of high efficiency power amplifiers. In *International Mobile Satellite Conference (IMSC)*, pages 59–64, Pasadena, California, June 1993.

[2] K.J. PARSONS, P.B. KENINGTON, AND J.P. MCGEEHAN. Efficient linearisation of RF power amplifiers for wideband applications. In *IEE colloquium on linear RF amplifiers and transmitters*, pages 7/1–7/7, London, England, April 1994.

[3] K.J. PARSONS AND P.B. KENINGTON. Effect of delay mismatch on a feedforward amplifier. *IEE Proceedings–G*, 141(2):140–144, April 1994.

[4] K.J. PARSONS AND P.B. KENINGTON. The efficiency of a feedforward amplifier with delay loss. *IEEE Transactions on Vehicular Technology*, 43(2):407–412, May 1994.

[5] K.J. PARSONS, P.B. KENINGTON, AND J.P. MCGEEHAN. High–efficiency power amplifier design for mobile satellite Earth stations. In *IEE colloquium on evolving technologies for small Earth station hardware*, pages 3/1–3/7, London, England, February 1995.

[6] P.B. KENINGTON, K.J. PARSONS, A. BATEMAN, AND J.P. MCGEEHAN. High efficiency power amplifier linearisation for mobile communications. In *Microwaves and RF Conference*, pages 24–27, Wembley, London, UK, October 1995.

[7] K.J. PARSONS, R.J. WILKINSON, AND P.B. KENINGTON. A highly–efficient linear amplifier for satellite and cellular applications. In *IEEE Global Telecommunications Conference*, pages 203–207, Singapore, November 1995.

ACKNOWLEDGEMENTS

Firstly, I should like to thank Prof. J.P. McGeehan for all of his advice throughout the last three years, and his provision of the excellent laboratory facilities which have been used in this work. I am also very grateful to Dr. Peter Kenington for his encouragement and technical support throughout my study.

For their financial support during my study, I am most appreciative to the Engineering and Physical Sciences Research Council (EPSRC) and Nortel (Europe) plc. In particular, I would like to thank Nigel King at Nortel for his enthusiasm, and thought–provoking questions during the project.

I am extremely grateful to my colleagues (and ex–colleagues) at the Centre for Communications Research, for all their comments and advice. However, there are one or two people who deserve special mention. Thanks must go to Ross Wilkinson, who, I'm sure everyone will agree is one of the most invaluable members of the CCR, and to Adrian Mansell, Wycliffe Slingsby, and Dave Bennett for their help on some of the more technical subjects. I am also grateful to Majid Boloorian for providing the computer code used to generate the $(\pi/4)$–DQPSK p.d.f.s in Chapter 5, and to Nigel Kingswood for sorting out all of my (and everyone else's) computer problems; where would we be without him? Thanks also to Mikey Fitton, for providing the lab with an endless source of entertainment, mostly just with his dress sense.

I would like to express my appreciation to Dr. Andrew Wathen of the Department of Mathematics for allowing me to use the optimal least–squares code in Chapter 6.

Although they have all been mentioned already, thanks are again due to Ross, Adrian, Wyc and Peter for their proof–reading of this work. It's a difficult job, and I appreciate their efforts greatly.

Finally, I am deeply indebted to all my family and friends for their support throughout my time in Bristol, without whom I could not have completed this work.

Chapter One

Introduction

1.1 Introduction

The demand for high quality radio links has increased dramatically over recent years. However, the amount of radio spectrum available for such applications has not increased by the same extent. This has led to an increasing demand for more spectrally efficient forms of radio communication. There has thus been a turn away from constant–envelope modulation towards linear modulation techniques. These are more spectrally efficient and potentially more power efficient than their constant–envelope counterparts. However, due to the envelope variations in the transmitted signal, intermodulation distortion will occur, which causes interference to both the same channel (co–channel interference) and those nearby in frequency (adjacent channel interference). To reduce the level of interference requires that the amount of intermodulation distortion be reduced: this can be achieved by using more linear transmitter (and receiver) amplifiers.

There is also increasing pressure to employ linear amplifiers in multi–carrier systems, for example in satellite transponders and cellular telephony basestations. Present terrestrial multi–channel systems generally employ individual channel amplifiers, with their outputs being combined at high level, using either hybrid or cavity combiners. Hybrid combiners [1] have the disadvantage that for each signal combination, half of the power is dissipated as heat. Their use is, therefore, highly power inefficient. Cavity combiners [2] are, in essence, very narrowband filters which select the wanted channel whilst providing a high impedance to all the other channels. The cavities must generally be manually retuned to allow for frequency allocation changes, and this also precludes the use of dynamic channel allocation. It is also not possible to use adjacent channels in the same basestation due to the finite Q of the cavities. A broadband linear amplifier, however, can be used to eliminate all of the above problems. The channels are instead combined at low power and the resulting signal amplified by a linear amplifier. This allows all of the channels to be used and has better power efficiency than the hybrid combiner technique since combination is at low power levels. It also supports dynamic frequency allocation, which can potentially improve the capacity of a cellular network by making better use of all the channels.

1.2 Linear Amplifiers for Satellite Applications

In mobile satellite systems the adjacent channel distortion requirements are much less stringent than for terrestrial basestations. This is due to the reduced dynamic range of the signals arriving at the mobile, compared with terrestrial systems. Hence, any distortion produced in the transmitter is likely to be below the receiver noise floor.

The number of channels in a satellite system can vary from tens to several hundred, and hence the systems tend to be very wideband. The efficiency of the amplification process is critical, since an increase in efficiency can lead to a reduction in the power supply requirements and the size of heatsinking required [3, 4]. Thus, a smaller area of solar panels is required and a smaller, lighter structure required to support them. Therefore, the overall spacecraft weight is significantly reduced leading to lower launch costs.

1.3 Linear Amplifiers for Land–Mobile Applications

1.3.1 Mobile Units

In contrast to satellite systems, there are much greater dynamic range problems in land–mobile systems. This arises because the path losses of the communications links vary widely, and it is possible for a weak wanted signal to be masked by the distortion produced by a strong signal on a nearby channel. The 'near–far effect' has severe consequences in cellular applications due to the large number of mobile transmitters, the high amount of frequency reuse, and the potentially large power difference between adjacent channels. Any distortion produced by a channel transmitter may spill into adjacent channels causing interference.

If a constant–envelope modulation format is used, the adjacent channel emissions can be kept very low, since no intermodulation distortion will be produced. However, linear modulation formats will produce a large amount of distortion when passed through a practical nonlinear amplifier. For adjacent channel operation, the linearity of the transmitter amplifier is required to be exceptionally good, typically with unwanted distortion products below −70dBc.

Mobile units for cellular applications are usually narrowband, single channel systems, operating with high power efficiency, in order to maximise battery life. Mobile units may also utilise wideband modulation, such as the Direct Sequence–Code Division Multiple Access (DS–CDMA) systems [5]. Although such systems are often considered to be constant–envelope, their use of filtering of the coding sidelobes introduces envelope variation, thus requiring linear amplification.

1.3.2 Basestation Units

The basestation units in a cellular system generally amplify a large number of channels over a broad band, typically 30MHz at 900MHz. The unwanted emissions requirement is also very stringent, for similar reasons as discussed in Section 1.3.1. If the channels are combined at low level, the resultant signal will have a varying envelope, regardless of the modulation format. Combination at low power allows for improved flexibility over the use of cavity combiners and increased power efficiency over the use of hybrid combiners. However, envelope variations of the resultant signal will cause distortion products to be produced when amplified, so it is necessary to use a highly linear amplifier in the transmitter.

The efficiency required for the basestation amplifier is not the most critical factor in its design, however, it should be made as high as possible to reduce the cost of the service.

1.4 Project Requirements

The preceding sections illustrate the pressing need for highly efficient linear amplifiers in both narrowband and wideband applications. There are a large number of techniques already applicable to the narrowband case, however the design of wideband linear amplifiers with high efficiency has not been extensively studied. Therefore, this work will focus on the study of techniques suitable for the wideband case.

1.5 Summary of Thesis

The modelling of nonlinearities, particularly with regard to RF amplifiers, has significant importance throughout this thesis. Hence, Chapter 2 introduces the effects of nonlinearities on signals and the basic techniques required to model them.

Chapter 3 describes some of the RF amplifier topologies, and presents methods to improve the linearity and/or power efficiency of the basic RF amplifier. The techniques which can be used for the required application are identified for further study.

The feedforward linearisation technique is discussed in Chapter 4, and its linearity and power efficiency characteristics are theoretically derived. This leads to the formulation of a number of methods which can be used to improve the performance of the feedforward linearisation technique. A practical feedforward amplifier is presented, the results from which are compared with the derived theoretical results.

In Chapter 5, the analogue predistortion technique is described in depth, and factors which affect its linearity and efficiency are addressed. The effect of the amplifier nonlinearity on the input signal statistics is determined. The theoretical average efficiency of predistortion amplifiers with various representative input signals, using both class A and C amplifiers, is also presented.

Chapter 6 examines the piecewise linear predistortion technique in detail. Several possible circuit topologies are investigated in simulation, illustrating the much improved linearity performance compared with conventional analogue predistortion techniques.

In Chapter 7 a summary of the thesis is given, and areas for future work are identified.

REFERENCES

[1] R.C. SHAPIRO. 900MHz trunked radio systems ferrite hybrid combining for closed space transmitters. In *IEEE 39th Vehicular Technology Conference*, pages 1/349–1/352, San Francisco, California, May 1989.

[2] S. KAZEMINEJAD, D.P. HOWSON, A. BAGHAI, AND G. HAMER. Cellular radio transmitter combiners for narrow co–channel spacing. In *IEE 5th International Conference on Mobile Radio and Personal Communications*, pages 57–59, Coventry, England, December 1989.

[3] M. KUMAR, J.C. WHARTENBY, AND H.J. WOLKSTEIN. Predistortion linearizer using GaAs dual–gate MESFET for TWTA and SSPA use in satellite transponders. *IEEE Transactions on Microwave Theory and Techniques*, MTT–33(12):1479–1488, December 1985.

[4] B.G. EVANS, editor. *Satellite Communication Systems.* Institution of Electrical Engineers, London, second edition, 1991.

[5] Qualcom Inc. *An overview of the application of CDMA to digital cellular systems and personal cellular networks*, May 1992. Document Number EX60–10010.

Chapter Two

Nonlinear Modelling

This chapter begins by examining the nature of the distortion produced by a nonlinear amplifier. Consideration is then given to how these nonlinearities may be described mathematically, and hence simulated. A number of modelling domains and techniques are explained, before choosing the methods to be used for the mathematical analyses and simulations later in this work.

2.1 Introduction

All practical amplifiers will exhibit a certain degree of nonlinearity, and it is necessary to understand how this can be modelled. Such knowledge would allow computer simulation of a nonlinear amplifier system, and provide insight into methods for linearity improvement. Although component–level modelling of amplifiers is possible, accurate modelling is normally prohibitively complex. Therefore, the type of amplifier modelling to be described is restricted to a 'black–box' approach. This type of modelling replaces the actual amplifier by a simulacrum which simply relates the input and output signals.

2.2 Effects of a Nonlinear Element

If a signal is applied to a nonlinear device, the distortion produced may be observed as additional frequency components in the output spectrum. It has been experimentally shown that if the applied signal has a constant amplitude (*e.g.*, a single tone), the output spectrum includes not only the original frequency components, but also additional components at frequencies at integer multiples of the original frequency components (including dc), termed *harmonic distortion.* The regions centred about each of these frequencies are referred to as *harmonic zones.* If the input signal has a time–varying envelope, further frequency components, in addition to those due to harmonic distortion are present. These distortion products are termed *intermodulation distortion* (IMD), and are found centred about the original frequencies, and each of the harmonics (as well as dc). Intermodulation products (IMPs) centred about the harmonics are referred to as sum–IMPs, since they occur at the sum of integer multiples of the input frequencies; those close to dc and the input frequencies are referred to as difference–IMPs, since they occur at the difference between integer multiples of the input frequencies.

A common signal used to determine the linearity of a device (particularly in terms of IMD) is the two–tone test signal. This signal consists of two sinusoids with (usually) the same amplitude, and a small frequency separation. The envelope of this signal takes the form of a rectified sinewave with a frequency twice that of the tone separation, and hence the device is exercised through the full range of power levels. A typical nonlinear response with a two–tone test input is shown in Figure 2.1, and shows both harmonic and intermodulation distortion.

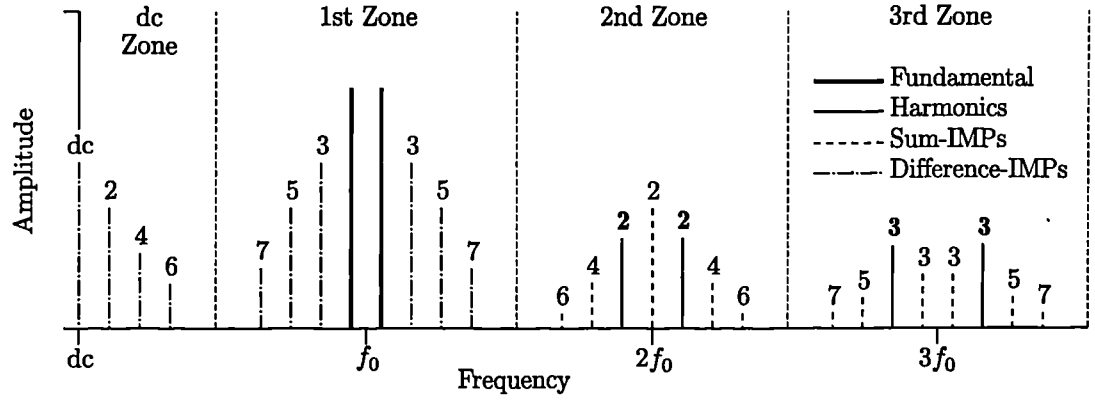


Figure 2.1: Spectrum produced by a nonlinear device (upto 7th order) with a two–tone input signal.

To investigate the effect of a nonlinearity on a constant–envelope signal, an input $x(t)$ con-

sisting of a single tone of amplitude A, and angular frequency ω,

$$x(t) = A\cos\omega t \tag{2.1}$$

is applied to a nonlinearity expressed as a third–degree polynomial given by[1]

$$y(t) = a_1x(t) + a_2x^2(t) + a_3x^3(t) \tag{2.2}$$

giving an output, $y(t)$, of

$$y(t) = \frac{1}{2}a_2A^2 + \left(a_1A + \frac{3}{4}a_3A^3\right)\cos\omega t + \frac{1}{2}a_2A^2\cos 2\omega t + \frac{1}{4}a_3A^3\cos 3\omega t \tag{2.3}$$

Thus, as expected, only harmonics are generated. It is important to note that the amplitude of the fundamental component is not only dependent upon the input amplitude but also upon the odd–degree terms of the nonlinearity (only third–degree in this case). Therefore, as the input amplitude is increased, the fundamental output amplitude rises nonlinearly. This gives rise to gain–expansion and gain–compression, which may be defined as those areas in the transfer characteristic where the incremental gain (defined as the gradient of the transfer characteristic) is greater than and less than 1dB per dB respectively.

To investigate the effect of the nonlinearity on an envelope–varying input signal, a two–tone test is used (which for simplicity has equal amplitude carriers), given by

$$x(t) = A(\cos\omega_1t + \cos\omega_2t) \tag{2.4}$$

Yielding an output

$$\begin{aligned}
y(t) &= a_2A^2 && \text{(dc)}\\
&+ \left(a_1A + \frac{9}{4}a_3A^3\right)(\cos\omega_1t + \cos\omega_2t) && \text{(Fundamental)}\\
&+ \frac{1}{2}a_2A^2(\cos 2\omega_1t + \cos 2\omega_2t) + \frac{1}{4}a_3A^3(\cos 3\omega_1t + \cos 3\omega_2t) && \text{(Harmonics)}\\
&+ a_2A^2\cos(\omega_1 - \omega_2)t + \frac{3a_3A^3}{4}\left(\cos(2\omega_1 - \omega_2)t + \cos(2\omega_2 - \omega_1)t\right) && \text{(Difference–IMPs)}\\
&+ a_2A^2\cos(\omega_1 + \omega_2)t + \frac{3a_3A^3}{4}\left(\cos(2\omega_1 + \omega_2)t + \cos(2\omega_2 + \omega_1)t\right) && \text{(Sum–IMPs)}
\end{aligned} \tag{2.5}$$

As would be expected, the output consists of both harmonic and intermodulation distortion. The *order* of the distortion is defined to be equal to the lowest *degree* term of the nonlinearity which generated it. This distinction is required because a k^{th}–degree nonlinearity generates distortion of order k or below, (with lower odd orders generated by odd–degree terms, and lower even orders generated by even–degree terms). Thus, the fundamental level is also dependent upon the third–degree term of the nonlinearity. If a higher degree polynomial was used to generate Equation 2.5, this effect would be seen on the IMPs and harmonics also.

In general, RF amplifiers have a bandwidth significantly less than the carrier frequency, and so bandpass filtering is used to reduce the harmonics to an acceptable level. However, when the input signal has a time–varying envelope the odd–order difference–IMPs produced in the first harmonic zone are too close to the fundamental signals to be filtered out, and this degrades performance in practical radio systems. Therefore, it is normally the case that only the first–zone output is required to be modelled.

[1]Note that at this stage it is not being stated that such a function can be used to model a nonlinear amplifier; it is merely being used as a simple analytical form to investigate the effects of a nonlinearity on a signal.

2.2.1 Asymmetry of Frequency Spectra

The frequency spectrum of the output of a nonlinear amplifier with a two–tone input often shows a large difference in amplitude between the same order of IMP on either side of the fundamental signals. The difference (up to 20dB) is too large to be due purely to the frequency–dependence of the nonlinearities of the amplifier. As a result, none of the nonlinear models to be presented are able to model this effect.

The physical causes of these effects are not completely known, and a number of suggestions have been published. Krauss *et al* [1] suggest that this is due to simultaneous amplitude and phase distortion. However this is incorrect, since memoryless nonlinear models (to be described in Section 2.4) can model simultaneous amplitude and phase distortion, but inherently produce symmetrical spectra. Pappenfus *et al* [2] have suggested that the asymmetry may be due to the interaction of IMPs when passing through a cascade of amplifiers. This can result in a large phase difference between IMPs generated by the first amplifier, and those generated by the second, and hence, can explain the large difference in IMP level on either side of the fundamental signals. However, it may be possible for the same effect to occur in a single stage amplifier, which cannot be easily explained by the above mechanism without a reasonable degree of feedback. It has been proposed by Sechi [3] that the asymmetry is due to the frequency response of the bias circuitry of the amplifier, and this is consistent with the fact that, in practice, the degree of asymmetry is generally reduced as the tone separation is reduced. However, none of these suggestions appears to completely explain the asymmetry of the frequency spectra, and, therefore, further study is required.

2.3 Modelling Domains

The nonlinearity can either be modelled in the time domain (relating the output signal at some time, t, with the input signal), the frequency domain (relating the output spectrum to the input spectrum), or a hybrid of the two.

2.3.1 Time Domain

Modelling in the time domain allows transient responses in the system to be analysed, with the relevant spectra being found using Fourier transforms. However, this method is inefficient for two main reasons. Firstly, the sampling rate must be high enough to avoid aliasing, although, as will be seen, this can be radically reduced by sampling the envelope of the signal, rather than the signal itself. Secondly, if energy–storage components (*e.g.*, filters) are present in the system, the number of samples required before the system has reached steady–state is excessive.

2.3.2 Frequency Domain

Frequency domain modelling allows efficient analysis of the steady–state conditions of a system, but it is not possible to determine its transient response. It is more difficult to model entirely in the frequency domain if nonlinearities are present, since new frequency components will be generated.

2.3.3 Hybrid Domain

Hybrid techniques exist (most notably *harmonic balance* [4–6]) which model the linear parts of the system in the frequency domain, and the nonlinear parts in the time domain, interfacing between the two using Fourier transforms. This makes the nonlinear modelling simpler than a purely frequency domain based modelling technique, whilst reducing the number of samples required compared with time domain techniques.

However, as with frequency domain techniques, only steady–state responses can be determined. The number of input frequencies which may be used with the harmonic balance method is limited, and the frequencies must generally be harmonically related.

2.4 Memoryless Bandpass Nonlinear Model

Models are often simplified by assuming that the nonlinearity does not exhibit memory, and thus the instantaneous output is only dependent upon the instantaneous input, and not previous inputs. The nonlinearity is, therefore, usually modelled by AM/AM and AM/PM conversion characteristics, which are measured by applying a constant envelope tone at the carrier frequency. The tone power is swept to provide the input to output power characteristic (AM/AM) and the phase shift versus input power characteristic (AM/PM) at the carrier frequency. The AM/PM conversion is, in fact, a memory effect since any phase shift requires knowledge of previous inputs; however, the memoryless models are usually modified to incorporate it.

By the definition of AM/AM and AM/PM conversion, it is apparent that only the first–zone characteristic is measured: *i.e.*, the output is measured through a bandpass filter which excludes components outside of the first zone. In the measurement process, it is the envelope level of a CW signal which is swept, producing the envelope transfer characteristic rather than the instantaneous RF voltage transfer characteristic. When modelling the system, the measured levels are treated as instantaneous power (defined as the square of the instantaneous envelope). It should also be noted that memoryless models can take no account of frequency–dependent variation of the system characteristics, which inherently requires memory.

In order to model a memoryless bandpass nonlinearity it is therefore necessary to use two nonlinearities operating on orthogonal versions of the input signal (to model both the AM/AM and AM/PM conversion). Thus, the nonlinearities operate on either the amplitude and phase (polar) components or quadrature (cartesian) components of the input signal.

2.4.1 Amplitude and Phase Representation

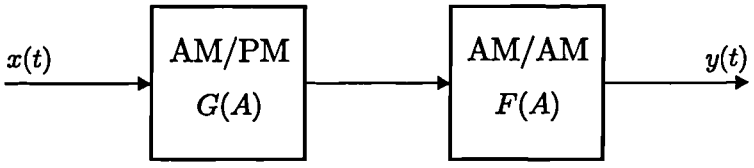


Figure 2.2: Amplitude and phase model of memoryless bandpass nonlinearity.

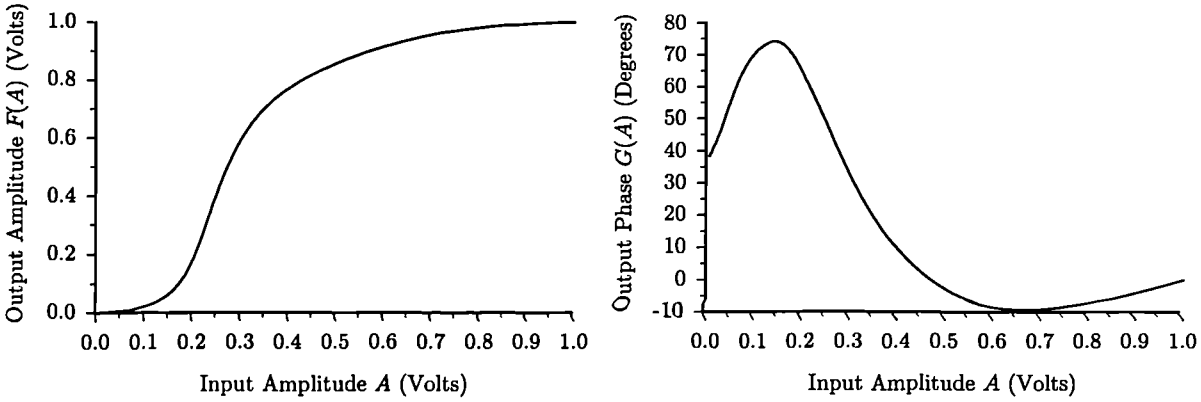
Consider a narrowband signal with amplitude and phase modulation, $x(t)$, thus

$$x(t) = A(t)\cos\left(\omega t + \phi(t)\right) \tag{2.6}$$

where $A(t)$ is the amplitude modulation and $\phi(t)$ is the phase modulation component. The output of a memoryless bandpass nonlinearity, $y(t)$, can be expressed as [2]

$$y(t) = F(A)\cos(\omega t + \phi + G(A)) \tag{2.7}$$

where $F(A)$ is the AM/AM conversion and $G(A)$ as the AM/PM conversion. This model is shown in Figure 2.2. The AM/AM and AM/PM conversion characteristics of a class C amplifier are shown in Figure 2.3. The class C amplifier characteristics presented throughout this chapter are measured using the VHF MOSFET class C amplifier to be described in Section 4.3.3 (page 66).



(a) AM/AM conversion

(b) AM/PM conversion

Figure 2.3: AM/AM and AM/PM conversion of a class C amplifier.

2.4.2 Quadrature Representation

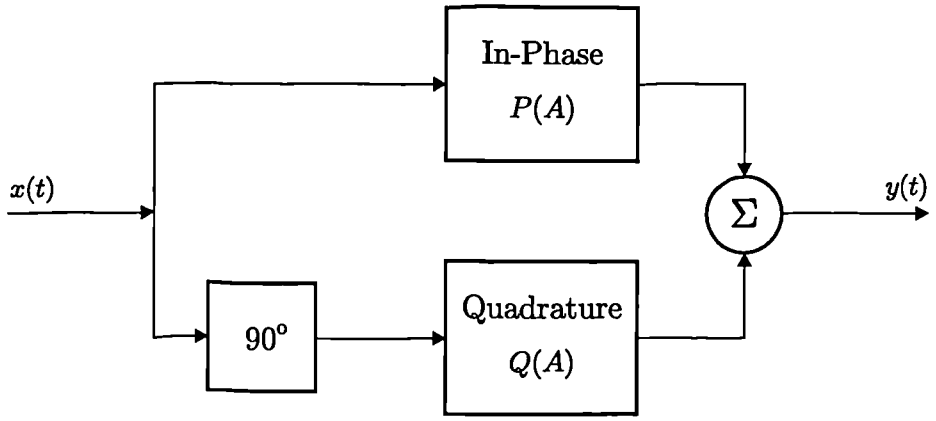


Figure 2.4: Quadrature model of memoryless bandpass nonlinearity.

The nonlinear output, $y(t)$, given by Equation 2.7 can be expanded to give

$$y(t) = F(A)\cos(G(A))\cos(\omega t + \phi) - F(A)\sin(G(A))\sin(\omega t + \phi) \tag{2.8}$$

[2]To avoid confusion between the envelope level, $A(t)$, and instantaneous signal level, $x(t)$, time will henceforth be made implicit for the modulation.

This can be expressed as the sum of two signals in quadrature, thus

$$y(t) = P(A)\cos(\omega t + \phi) - Q(A)\sin(\omega t + \phi) \tag{2.9}$$

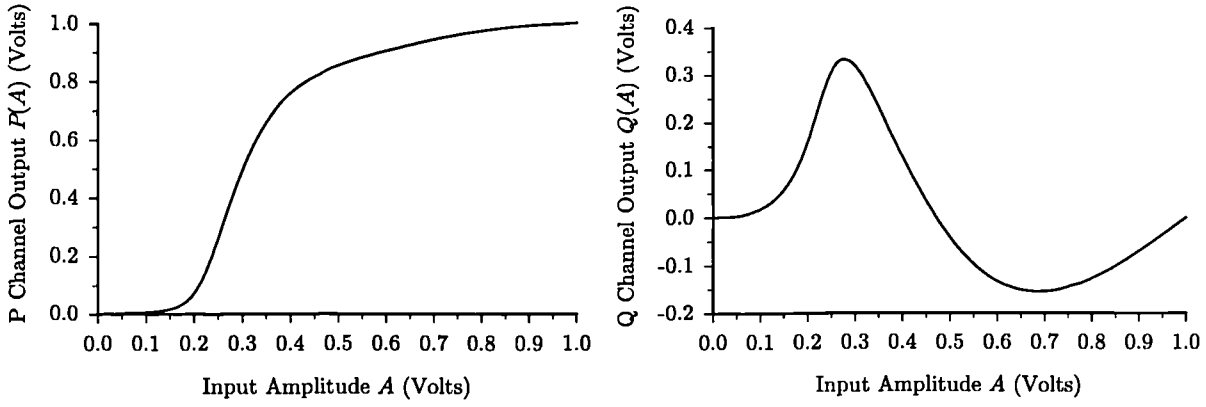
where

$$P(A) = F(A)\cos(G(A)) \tag{2.10}$$

and

$$Q(A) = F(A)\sin(G(A)) \tag{2.11}$$

This model is shown in Figure 2.4. It can be seen, therefore, that the AM/AM and AM/PM conversion can be modelled by two nonlinearities which have only AM/AM conversion. The quadrature form of the AM/AM and AM/PM characteristics of the class C amplifier (Figure 2.3), are shown in Figure 2.5.



(a) In–phase component

(b) Quadrature component

Figure 2.5: Quadrature equivalent of AM/AM and AM/PM conversion of a class C amplifier.

It can easily be shown that if the degree of AM/PM conversion is small (*i.e.*, $G(A) \approx 0$), the following approximations of the in–phase and quadrature nonlinear functions can be made.

$$P(A) \approx F(A) \tag{2.12}$$

$$Q(A) \approx F(A)G(A) \tag{2.13}$$

2.4.3 Complex Envelope Representation

The analytical representation of a modulated carrier is given by

$$x(t) = A(t)e^{j(\omega_0 t + \phi(t))} \tag{2.14}$$

where $A(t)$ is the amplitude modulation and $\phi(t)$ is the phase modulation. This may be written as

$$x(t) = A(t)e^{j\phi(t)}e^{j\omega_0 t} \tag{2.15}$$

The information present in the modulation is completely contained in the signal $r(t)$, known as the complex envelope, given by

$$r(t) = A(t)e^{\jmath\phi(t)} \tag{2.16}$$

Hence

$$x(t) = r(t)e^{\jmath\omega_0 t} \tag{2.17}$$

It is possible to show that low–pass filtering of the complex envelope, $r(t)$, is equivalent to band–pass filtering of the RF signal, $x(t)$. Hence, if the RF signal has a bandwidth B, the frequency of the complex envelope extends from dc to $B/2$, and hence a sampling rate of at least B would be required to avoid aliasing. However, to sample the RF signal, $x(t)$, without aliasing requires a sampling frequency greater than $\left(\frac{\omega_0}{\pi} + B\right)$. Hence, the use of the complex envelope can significantly reduce the sampling rate required, and thus decrease simulation time.

The RF signal $x(t)$ can be written in an equivalent form as

$$x(t) = A(t)[\cos(\omega_0 t + \phi(t)) + \jmath \sin(\omega_0 t + \phi(t))] \tag{2.18}$$

This representation of a signal has a number of advantages. Firstly, to phase shift the signal by θ, all that is required is to multiply the signal by $e^{\jmath\theta}$ (equivalent to $\cos\theta + \jmath\sin\theta$). Secondly, the envelope of the signal is simply equal to its magnitude.

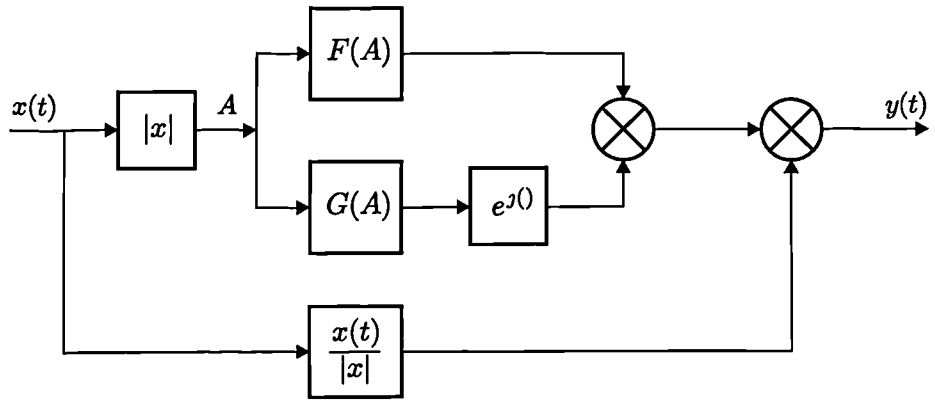
The memoryless bandpass nonlinear models derived in Sections 2.4.1 and 2.4.2 can be adapted to use the complex envelope of the input signal, resulting in efficient simulation of the nonlinear device. The memoryless bandpass nonlinear models using amplitude and phase, and quadrature representations (Figures 2.2 and 2.4), are shown in Appendix A to have complex envelope equivalents as illustrated in Figure 2.6.

2.5 Envelope and Instantaneous Nonlinear Models

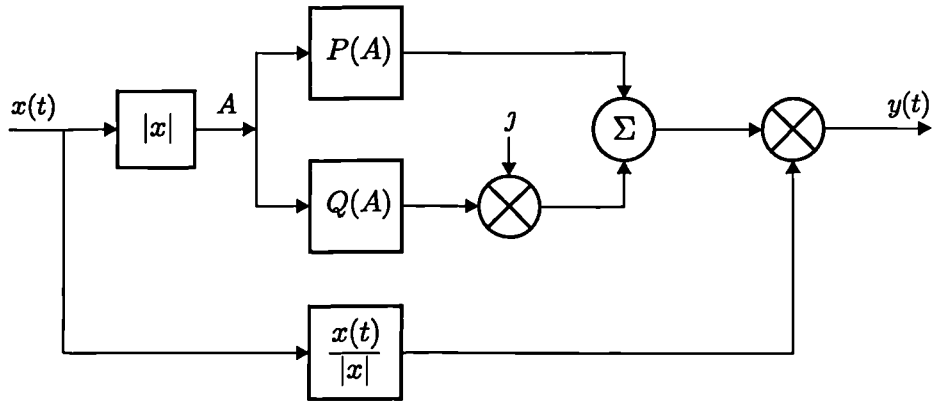
The memoryless models previously described relate the output envelope and phase to the input envelope, giving rise to envelope transfer characteristics. However, it is often useful to consider how the instantaneous output voltage is related to the instantaneous input voltage (*i.e.*, the instantaneous transfer characteristic).

An instantaneous transfer characteristic is usually found by applying dc or a ramp signal to the device, which is clearly not possible at RF, due to filtering effects. However, it is possible to convert the envelope transfer characteristic to an instantaneous transfer characteristic and vice versa, using the method to be described. Note that the envelope characteristic cannot completely determine the nonlinearity (since only the first–zone response is determined), and thus the actual instantaneous transfer characteristic may not be found from it. However, the two forms of the nonlinear function can be equated to give the same level of nonlinearity in the first–zone.

The first–order Chebyshev transform [7] (also known as the describing function [8]) can be used to transform a function from the instantaneous form to the envelope form. It is also possible to invert the first–order Chebyshev transform to allow transformation from the envelope form to the instantaneous form, as discussed in [7].



(a) Amplitude and phase representation



(b) Quadrature representation

Figure 2.6: Complex envelope equivalent models of a bandpass memoryless nonlinearity.

2.5.1 Chebyshev Transform

Consider a narrowband signal with amplitude and phase modulation, $x(t)$, thus

$$x(t) = A(t)\cos\left(\omega t + \phi(t)\right) \tag{2.19}$$

and make the substitution

$$\alpha(t) = \omega t + \phi(t) \tag{2.20}$$

The output $y(t)$ of an instantaneous nonlinearity, $f(x(t))$, may therefore be written as a function of $\alpha(t)$

$$y(t) = f(A(t)\cos\alpha(t)) \tag{2.21}$$

This is an even periodic function, and can be expanded as a Fourier series (making t implicit)

$$y = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos k\alpha + b_k\sin k\alpha) \tag{2.22}$$

The first–zone output is given with $k = 1$, thus

$$y = a_1\cos\alpha + b_1\sin\alpha \tag{2.23}$$

where

$$a_1 \equiv C_1(A) = \frac{1}{\pi}\int_0^{2\pi} f(A\cos\alpha)\cos\alpha\, d\alpha \tag{2.24}$$

$$b_1 \equiv C_2(A) = \frac{1}{\pi}\int_0^{2\pi} f(A\cos\alpha)\sin\alpha\, d\alpha \tag{2.25}$$

Equation 2.24 has been termed the first–order Chebyshev transform by Blachman [7], and in control theory Equations 2.24 and 2.25 are combined to form the quantity known as the describing function [8,9], given by

$$N = \frac{C_1(A)}{A} + j\frac{C_2(A)}{A} \tag{2.26}$$

The output of the envelope nonlinearity is therefore

$$y(t) = C_1(A)\cos(\omega t + \phi(t)) + C_2(A)\sin(\omega t + \phi(t)) \tag{2.27}$$

It can be seen by comparison of Equations 2.9 and 2.27 that

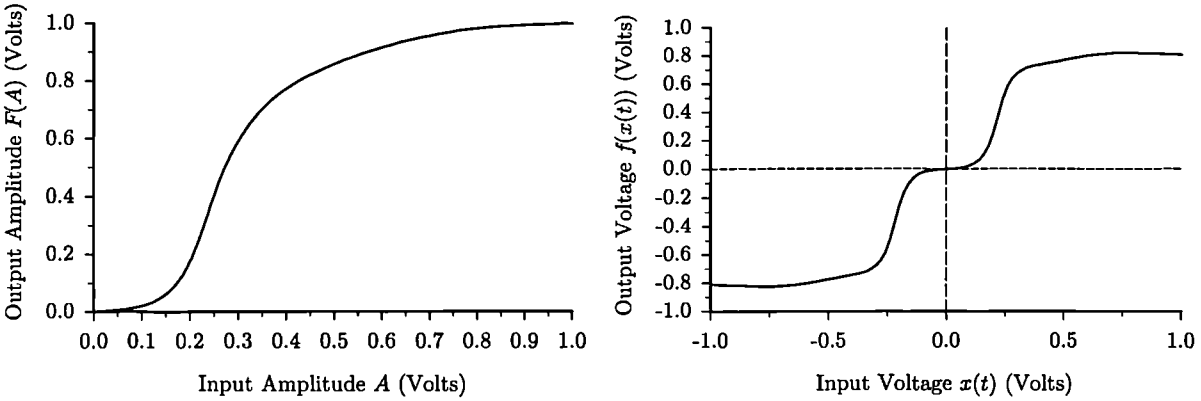
$$C_1(A) \equiv P(A) \tag{2.28}$$

$$C_2(A) \equiv -Q(A) \tag{2.29}$$

The envelope characteristic of a nonlinearity with purely AM/AM conversion can be transformed into its instantaneous counterpart by using the inverse Chebyshev transform [7].

$$f(x(t)) = \frac{1}{2}\int_0^{\frac{\pi}{2}} \left\{C_1(x(t)\cos\phi(t)) + x(t)C_1'(x(t)\cos\phi(t))\right\} d\phi \tag{2.30}$$

The inverse of an envelope nonlinearity with both AM/AM and AM/PM conversion cannot be readily determined since the instantaneous function is no longer single valued. The envelope and instantaneous nonlinear transfer functions for the aforementioned class C amplifier (assuming AM/AM conversion only) are shown in Figure 2.7.



(a) Envelope

(b) Instantaneous

Figure 2.7: Envelope and instantaneous transfer characteristics of class C amplifier.

2.6 Optimal Transfer Characteristic of an Amplifier

An optimal transfer characteristic can be conceived, with which the generation of first zone distortion components is minimal. This mathematical abstraction, whilst being unobtainable in practice, would allow the response of a practical amplifier to be compared with that of the ideal amplifier. Such a transfer characteristic would have no gain or phase variation with varying input envelope. However, given that the amplifier must saturate at some level, this is an unreasonable aim.

It has been suggested in [10] that, given that the amplifier must eventually saturate, the ideal characteristic would be that of a piecewise linear limiter. This has an amplitude transfer characteristic which increases linearly with input magnitude up to some value, K, and then instantly limits, as shown in Figure 2.8. The phase characteristic has zero phase, independent of input magnitude.

$$F(A) = \begin{cases} A, & A \leq A_0 \\ K, & A > A_0 \end{cases} \tag{2.31}$$

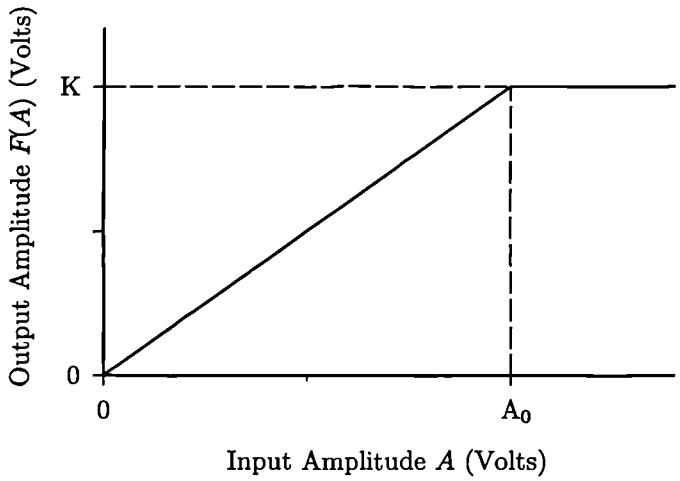


Figure 2.8: Amplitude Transfer characteristic of a piecewise limiter.

However, whilst this may intuitively appear to be the optimal characteristic, it has been shown in [11] that it can be suboptimal under certain conditions, which are dependent upon the input waveform. However, it is impractical to alter the characteristic as the input waveform changes, and so this is usually used as the design goal.

2.7 Approximations to Memoryless Nonlinear Model

The nonlinear functions used in the memoryless model may be approximated by using a number of methods, described in the following sections. There are several properties which are desirable in the approximating function. The function should be mathematically tractable to allow algebraic manipulation and/or be efficient to compute. It should also be able to model practical nonlinear systems, either by approximating actual measured data, or by approximating the general shape of the nonlinearity.

2.7.1 Determination of the approximating function

The nonlinear function, $f(x)$ is measured at a finite number of points, x_j, where $j = 1, \ldots, n$, which will have a certain error. The approximation goal is to find a function, $\zeta(x)$, using the measured data, which provides the 'best' approximation to the actual nonlinear function, $f(x)$. Denoting the true value of $f(x_j)$ as f_j, and the measured value as $\bar{f}_j$, the error, E_j, is given by $E_j = f_j - \bar{f}_j$. It is normally assumed that the abscissæ are free of error, in order to reduce the difficulty in finding the approximating function to a manageable level.

The error is made up of two components; *statistical errors* (*e.g.*, noise), which can be reduced by averaging the data samples, and *systematic errors* (*e.g.*, calibration inaccuracies) which cannot be reduced by averaging [12].

Once an approximating function has been selected, its form is chosen to fit the measured data in some way. The underlying function which the data represents is assumed to be smooth, *i.e.*, has low average curvature, and so the smoother the approximating function, for a given error, the better the approximation is deemed to be.

Interpolation

Interpolation [12] requires the approximating function, $\zeta(x)$, to pass through all the measured data points, and hence no account is taken of any errors in the data. The function chosen is usually a polynomial, and interpolation can either be performed in a local manner, where a number of piecewise polynomials are used, or a global manner in which a single polynomial is used.

If a single polynomial is to be used, its degree is required to be equal to $n - 1$, and thus the number of terms to be calculated is generally excessive. Polynomials also have the property that as the degree is increased, its smoothness is reduced, making its use less desirable.

To reduce these effects it is common to use a piecewise fit of a number of low degree polynomials in preference to a single high degree one. These functions are termed *splines*, with the most common using a third–degree polynomial, known as the cubic spline, which has the properties of smooth first derivative and continuous second derivative at the interface of the piecewise functions. It may be shown [13] that if the second derivatives at the first and last data points is set to zero (the so–called natural spline), the cubic spline fit is the smoothest exact–matching fit between the data points. One disadvantage of using splines is that they are not easily algebraically manipulated over the domain of the data.

Least–squared approximation

To determine the coefficients of the approximating function a least–squared error fitting procedure is commonly used [14]. This is achieved by choosing the coefficients so as to minimise

$$\sum_{j=1}^{n} w(x_j)\left[\bar{f}_j - \zeta(x_j)\right]^2 \tag{2.32}$$

where $w(x)$ is a weight function, assumed to be such that $w(x_j) \geq 0, j = 1, \ldots, n$. Often $w(x_j)$ is taken to equal one.

The approximating function is no longer constrained to pass through the data points, but instead provides a fit which is based on minimising the error in some global manner. It can be shown that if the data points have a measurement error which is independently random

and has a normal distribution, the least–squared approximation is a maximum likelihood estimator, and thus produces an optimal fit. Whilst this may be the case with statistical errors, it is not so with systematic errors, and thus the approximation is not generally optimal. However, its easy mathematical tractability makes the method popular even when its use is not necessarily applicable.

Minimax approximation

The Minimax approximation [14] produces a fit which minimises the maximum error between the data points and the approximating function, thus

$$\max_{j=1,\ldots,n} w(x_i)|\bar{f}_j - \zeta(x_j)| \tag{2.33}$$

The major disadvantage of this method is that the coefficients of the approximating function are very difficult to find.

2.7.2 Polynomial Approximation

The Weierstrass Approximation Theorem [13] states that if f is defined and continuous on the interval $[a, b]$, and $\varepsilon > 0$ is given, then there exists a polynomial, $\mathcal{P}$, with the property that

$$|f(x) - \mathcal{P}(x)| < \varepsilon \quad \text{for all} \quad x \in [a, b] \tag{2.34}$$

Thus there exists a polynomial such that the error between the function and the polynomial approximation (in the minimax sense at least) can be made arbitrarily small. (Note that the order of the polynomial required is not given.)

The approximating function may be given as a linear combination of weighted polynomials [15, 16] thus

$$\zeta(x) = \sum_{i=0}^{n} a_i g_i(x) \tag{2.35}$$

where a_i are real constants, and $g_i(x)$ are polynomials.

The polynomials, $g_i(x)$, which can be used in the series given in Equation 2.35, are many and varied. The following sections describe some of the most used polynomials.

Power Series

A polynomial may be written as a power series, using $g_i(x) = x^i$, thus

$$\zeta(x) = \sum_{i=0}^{n} a_i x^i \tag{2.36}$$

where a_i are real constants.

Note that all polynomials can be represented in this form, however, in general this conversion should not be done. In many approximations the coefficients a_i do not decrease with degree, and the series is termed ill–conditioned, causing large inaccuracies with finite–precision computation. Another polynomial series may have well–conditioned coefficients, and hence can be calculated more accurately.

A useful feature of using this type of polynomial is that the magnitude of the first–zone IMPs may be calculated directly from the polynomial coefficients [17].

Chebyshev Approximation

The Chebyshev polynomial [18], $T_n(x)$, of degree n is given by

$$T_n(x) = \cos\left(n \cos^{-1} x\right) \tag{2.37}$$

which is given explicitly by a polynomial

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ &\vdots \\ T_{n+1} &= 2xT_n(x) - T_{n-1}(x) \quad n \geq 1 \end{aligned} \tag{2.38}$$

The approximating function $\zeta(x)$ can be given by a sum of Chebyshev polynomials of differing degrees, thus

$$\zeta(x) = \frac{1}{2}c_0 + \left[\sum_{k=1}^{n-1} c_k T_k(x)\right] \tag{2.39}$$

where n is the maximum Chebyshev polynomial degree, and c_k are real constants.

The Chebyshev polynomial coefficients may be found using least–squared algorithms; alternative algorithms allow the Chebyshev polynomial to approximate the Minimax polynomial (but only if a particular set of abscissæ is used, which may not be the case with measured data).

It can be shown that if the approximating function given in Equation 2.39 is truncated the error introduced can be no larger than the sum of the neglected c_k's. Since the series is well–conditioned, the coefficients, c_k, become smaller as the polynomial degree is increased. Thus, the error introduced is generally small and also tends to be smoothly spread throughout the range of the function.

Orthogonal Functions

A set of functions $\{f_0, f_1, \ldots, f_n\}$ is *orthogonal* in the interval $[a, b]$ with respect to the weight function, w, if

$$\int_a^b w(x) f_j(x) f_k(x)\, dx = \begin{cases} 0 & \text{when } j \neq k \\ \alpha_k > 0 & \text{when } j = k \end{cases} \tag{2.40}$$

Functions that are orthogonal include Fourier series and Bessel functions, and also many polynomials such as Chebyshev, Legendre and Hermite polynomials [19].

The use of a series of orthogonal functions for approximation has a number of very useful advantages. In particular, the coefficients of the expansion may be calculated separately, rather than having to solve a number of linear simultaneous equations, *i.e.*, the computational labour is of the order n rather than n^3 [13].

2.7.3 Rational Functions

The approximating function, $\zeta(x)$ can also be given by a rational function (*i.e.*, quotients of polynomials) [14]

$$\zeta(x) = \frac{\sum_{i=0}^{n} a_i g_i(x)}{\sum_{i=0}^{m} b_i g_i(x)} \tag{2.41}$$

It is usual to use rational functions which have similar degree numerators and denominators, since in this case the approximation accuracy is generally superior to the polynomial case for the same amount of computational expense.

Unlike polynomial approximation, rational functions (and continued fractions) have the property that they produce an infinite number of first–zone IMPs in the output spectrum. Whilst this makes it difficult to ascertain the magnitude of a particular product from the approximating function coefficients, it does allow an algebraically simple function to model the effects of a large number of IMPs (using polynomials the maximum order of IMPs generated is equal to the highest degree term in the polynomial).

Saleh Functions

It has been suggested by Saleh [20] that relatively simple rational functions can be used to approximate the nonlinear characteristics of amplifiers. The proposed method uses two–parameter formulæ to represent the AM/AM and AM/PM conversion directly, or to represent the quadratic model derived in Section 2.4.2. Using the same terminology as previously, Saleh approximates the AM/AM and AM/PM conversion by

$$F(A) = \frac{\alpha_a A}{(1 + \beta_a A^2)} \tag{2.42}$$

and

$$G(A) = \frac{\alpha_\phi A^2}{(1 + \beta_\phi A^2)} \tag{2.43}$$

To represent the quadratic model, the following two–parameter formulæ have been proposed

$$P(A) = \frac{\alpha_p A}{(1 + \beta_p A^2)} \tag{2.44}$$

and

$$Q(A) = \frac{\alpha_q A^3}{(1 + \beta_q A^2)^2} \tag{2.45}$$

The required parameters (α and β) are chosen using a least–square error curve fitting procedure to the measured AM/AM and AM/PM characteristics. The formulæ were designed specifically for TWT characteristic modelling, although other saturating type characteristics (such as class A or class AB amplifiers) may be accurately modelled. However, highly nonlinear characteristics, such as those of class C amplifiers are not modelled with a satisfactory degree of accuracy.

2.7.4 Continued Fractions

The approximating function, $\zeta(x)$, may also be written as a continued fraction [14], given by

$$\zeta(x) = \cfrac{1}{1+\cfrac{c_1 x}{1+\cfrac{c_2 x}{1+\cdots+\cfrac{c_{2n-1}x}{1+c_{2n}x}}}} \tag{2.46}$$

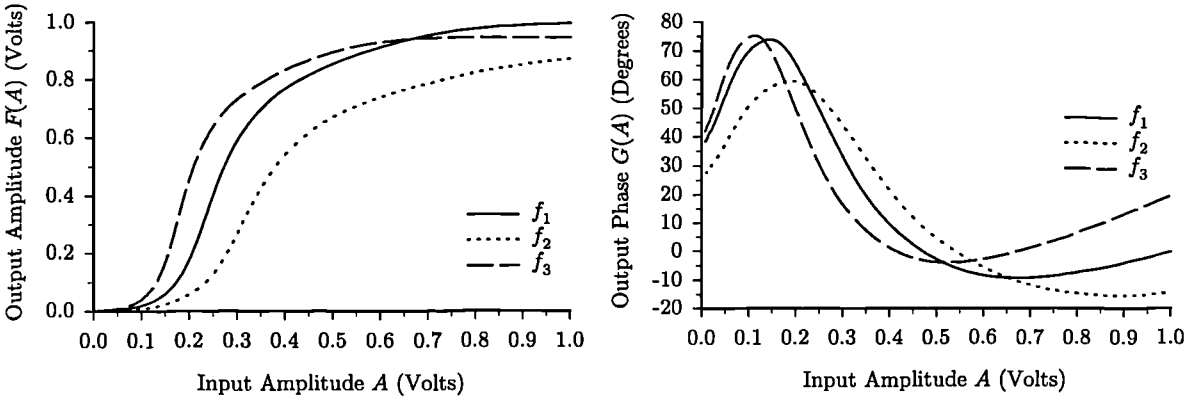
It is possible to convert rational functions to and from continued fractions. However, the continued fraction is a more computationally–efficient representation than the rational function.

2.7.5 Miscellaneous Functions

There are a number of other functions which can be used to approximate the required nonlinear functions which will be mentioned for completeness. These include Fourier series [21], Bessel functions [22], Barrett–Lampard expansions [23], and a number of customised functions, such as those proposed by Shimbo [24]. These functions are, however, generally less efficient or more complex than those previously described.

2.8 Bandpass Nonlinear Model with Memory

The memoryless model of a nonlinear device allows reasonable modelling of an amplifier. However, although it can be extended to allow for the effects of AM/PM conversion, it cannot be readily extended to allow for other memory effects. The most important of these in communications systems is the frequency–dependence of the amplifier characteristic. This causes the AM/AM and AM/PM conversion characteristics to alter with frequency, and these are shown for a class C amplifier in Figure 2.9.



(a) AM/AM conversion

(b) AM/PM conversion

Figure 2.9: AM/AM and AM/PM conversion of a class C amplifier at various frequencies.

There are two methods generally used to model an amplifier nonlinearity with memory, the so–called *classical* and *basis function* approaches. Recently, an alternative method has been proposed which has the potential for higher degrees of accuracy in the model.

2.8.1 Classical Approach

This approach to nonlinear modelling incorporating memory effects uses an isolated memoryless nonlinearity, with input and output linear filters, as shown in Figure 2.10. The filters provide the memory in the system, and allow for a frequency–dependent characteristic.

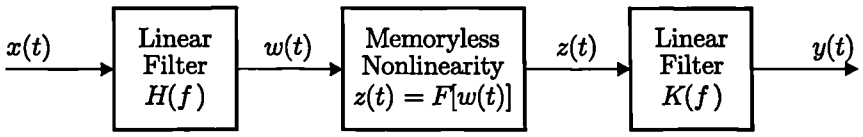


Figure 2.10: Classical model of a nonlinearity with memory.

Any of the methods used to approximate memoryless nonlinear models described in Section 2.7 can be used to model the memoryless nonlinearity in the classical model. However, some more specific methods have been designed, which are described in the following sections.

There have been a number of attempts to derive a nonlinear model with memory based on the quadrature representation of a memoryless bandpass nonlinearity (Section 2.4.2). These models are based on making the in–phase and quadrature nonlinearities frequency–dependent.

Saleh Model

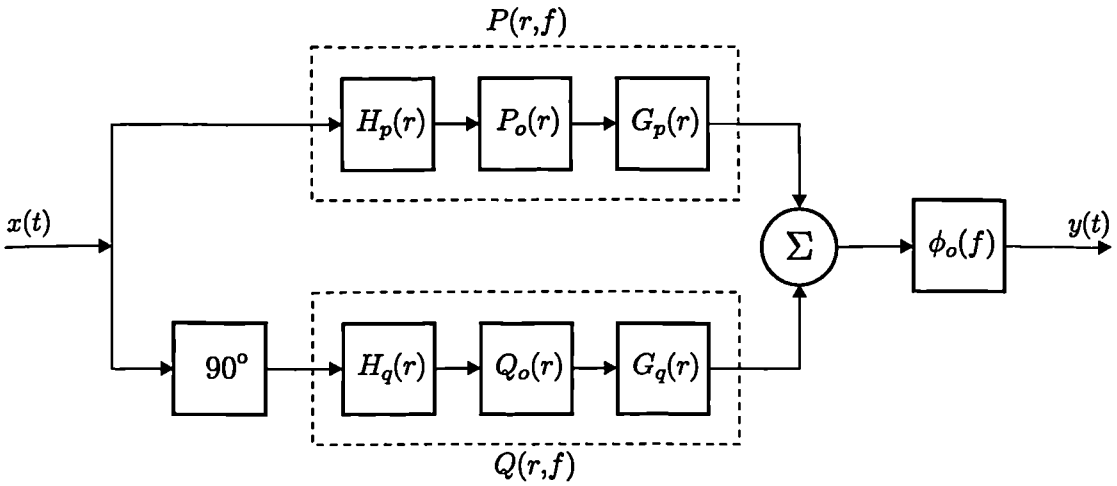


Figure 2.11: Saleh frequency–dependent quadrature model.

The functions described in Section 2.7.3 have been extended in [20] to include frequency–dependence of the nonlinear functions by allowing the coefficients α and β in Equations 2.42 to 2.45 to be frequency–dependent. This has been shown to be equivalent to adding real filters before and after the in–phase and quadrature nonlinearities, and results in the block diagram configuration shown in Figure 2.11, thus fulfilling the requirements of the classical approach.

It has been shown in [25] that the model requires the shape of the in–phase and quadrature nonlinearities to be independent of frequency, with only a scaling factor providing the frequency–dependence. Whilst the general shape of the nonlinearities will not change by a large degree, the inability to model this effect will reduce the accuracy which can be attained using this model.

Abuelma'atti Model

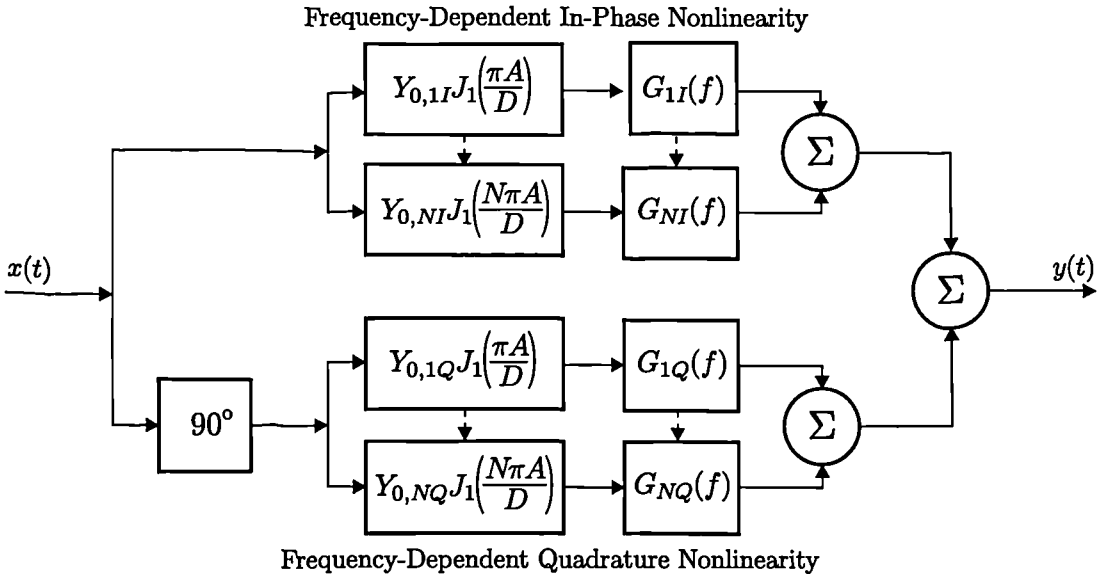


Figure 2.12: Abuelma'atti frequency–dependent quadrature model.

A similar, but more complex, frequency–dependent quadrature model has been proposed by Abuelma'atti [26]. The in–phase and quadrature nonlinearities have been expanded as the sum of first–order Bessel functions, followed by real filters to provide the frequency–dependence, as shown in Figure 2.12. The model allows for the shape of the characteristics to be different, and therefore may provide more accurate modelling than the frequency–dependent Saleh functions. This may be considered as an extension to the classical approach, since the shape of the nonlinearity is no longer required to be constant.

2.8.2 Basis Function Approach

The basis function approach is an extension of the classical approach, in which the separation of the linear and nonlinear parts of the system is removed. The input, $x(t)$, is split between N nonlinear blocks (basis functions), and then summed to give the output, $y(t)$, as shown in Figure 2.13.

The most common methods using the basis function approach are Volterra series, and Generalised Power series.

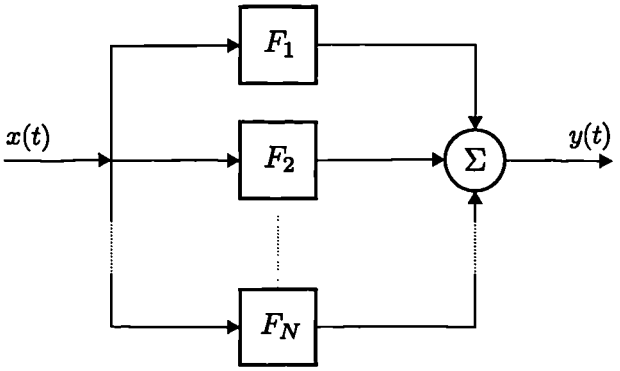


Figure 2.13: Basis function model of a nonlinearity with memory.

Volterra Series

The output of a linear time–invariant system is given by the convolution integral.

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,d\tau \tag{2.47}$$

where $h(\tau)$ is the impulse response of the system. In an analogous manner, it was suggested by Wiener [27] that the output of a (weakly) nonlinear system, $y(t)$, may be expressed as a series of Volterra kernels

$$\begin{aligned} y(t) = & \int_{-\infty}^{\infty} h_1(\tau_1)x(t-\tau_1)\,d\tau_1 \\ & + \int_{-\infty}^{\infty}\int h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)\,d\tau_1\,d\tau_2 \\ & + \int_{-\infty}^{\infty}\iint h_3(\tau_1,\tau_2,\tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)\,d\tau_1\,d\tau_2\,d\tau_3 \\ & + \ldots \end{aligned} \tag{2.48}$$

This can be expressed more compactly as

$$y(t) = \sum_{n=1}^{\infty} y_n(t) \tag{2.49}$$

where

$$y_n(t) = \int_{-\infty}^{\infty}\cdots\int h_n(\tau_1,\ldots,\tau_n)\prod_{i=1}^{n} x(t-\tau_i)\,d\tau_i \tag{2.50}$$

The function $h_n(\tau_1,\ldots,\tau_n)$ is the n^{th}–order Volterra kernel of the system, and is generally known as the n^{th}–order nonlinear impulse response.

The time–domain form of the Volterra series given above can be converted to a frequency–domain version (which is more readily used), thus

$$y_n(t) = \int_{-\infty}^{\infty}\cdots\int H_n(\omega_1,\ldots,\omega_n)\prod_{i=1}^{n} X(\omega_i)e^{j\omega_i t}\,d\omega_i \tag{2.51}$$

where $H_n(\omega_1,\ldots,\omega_n)$ is the n^{th}–order nonlinear transfer function of the system, and $X(\omega)$ is the Fourier transform of the input signal. In this form the two linear transfer functions, $H(\omega)$ and $K(\omega)$, present in the classical approach have been replaced by n nonlinear transfer functions, $H(\omega_1)$ to $H(\omega_1,\omega_2,\ldots,\omega_n)$.

The Volterra series has been used extensively to model nonlinearities with memory [28–34], but has several limitations; the convergence of the model is slow, and the determination of higher–order terms from measured data is difficult. However, as with polynomial approximation, it is possible to calculate the effect which individual orders of nonlinearity have on the intermodulation distortion (or harmonic distortion).

Generalised Power Series

The generalised power series [35,36] is a much more efficient method of describing a nonlinearity with memory than the Volterra series, and can handle severe nonlinearities.

For a system with a multi–frequency input, $x(t)$, given by

$$x(t) = \sum_{i=1}^{N} x_i(t) = \sum_{i=1}^{N} |x_i| \cos\{\omega_i t + \phi_i\} \tag{2.52}$$

the output, $y(t)$, is described by the generalised power series

$$y(t) = A \sum_{k=0}^{\infty} a_k \left\{ \sum_{i=1}^{N} b_i x_i(t - \tau_{i,k}) \right\}^k \tag{2.53}$$

where k is the order of the power series; a_k is a complex coefficient; b_i is a real coefficient; and $\tau_{i,k}$ is a time delay that is dependent upon frequency and the power series order. The use of complex coefficients and frequency–dependent time delays allows a broad class of nonlinear systems with memory to be analysed. It has been shown [37] that there is a relationship between the Volterra and generalised power series. However, it is not clear how to generate the required coefficients from measured device data.

2.8.3 Blum and Jeruchim Model

This model does not fall into either of the previous categories, and appears to be a unique method. In the models thus far presented, the amplifier characteristics have been measured using a power–swept sinewave, and altering the frequency to generate a series of curves. It is stated by Blum and Jeruchim [25] that this cannot provide all the information required to completely model the amplifier, since it cannot allow for the interaction of the frequency components of a wideband signal when passed through a nonlinear device. Therefore, they suggest that the use of a number of equally spaced tones, generated by multiplying an RF carrier by a pseudo–random (PN) sequence, would allow more accurate modelling. This signal is applied to the nonlinear device, and the output is passed through a narrow bandpass filter to measure the amplitude of a particular frequency component. The centre frequency of the filter is swept across the band, and a number of characteristics, with different input powers, are measured. The PN–sequence measures the frequency–dependence of the device at a particular input power since it is constant–envelope, and the changing input power allows the nonlinear effects to be measured.

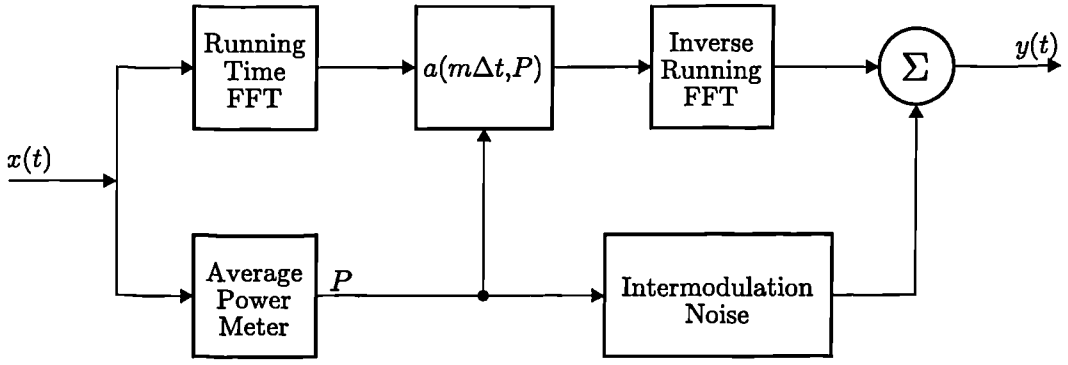


Figure 2.14: Blum and Jeruchim nonlinear model with memory.

To model the nonlinearity, the input signal is converted into the frequency domain using a fast–Fourier transform (FFT), and passed through a power–dependent transfer function. This signal is then converted back into the time domain using an inverse FFT. However, whilst the model will allow for the influence of different frequency tones in the input signal on a particular output tone, it has been acknowledged that its treatment of IMPs may not be exactly correct. This has resulted in the use of an extra block in the model to generate additional IMPs. The resulting model is shown in Figure 2.14. In its current form only the amplitude of a particular frequency component is measured, and so the model cannot take account of the AM/PM conversion. Thus far there have been no comparisons between the model and experimental results published and, therefore, the accuracy of the model, whilst potentially good, cannot be ascertained.

2.9 Summary

This chapter describes the measurable effects of distortion, in terms of harmonic and intermodulation signal components. Although the physical mechanisms which cause this distortion are highly complex, they may be approximated by a number of mathematical models. The use of these models in computer simulations of nonlinear amplifiers permit a better understanding of their operation, and ultimately will allow various linearisation strategies to be modelled.

The complex envelope notation has been introduced, which allows efficient modelling of nonlinear devices, due to the lack of harmonics in the output spectrum, and the ease of calculating the signal envelope.

A number of different models have been examined in this chapter, both incorporating and excluding the memory effect of the nonlinear device. The memoryless models are less complex, and have simpler measurement methods, but cannot incorporate the effects of the frequency–dependence of the nonlinearity, and so are restricted to the narrowband case only.

Many of the techniques are only suitable for weakly nonlinear systems, due to their complexity. The bandpass memoryless nonlinear model will be used in later chapters, due to its ability to model highly nonlinear systems with easily measurable characteristics. Polynomial approximation will be used to analytically investigate amplifier system responses, in preference to the other techniques, due to its relatively simple mathematical tractability. In the computer simulation work to be presented, interpolation of measured data allows accurate modelling of the highly nonlinear nature of the amplifiers used.

REFERENCES

[1] H. L. KRAUSS, C. W. BOSTIAN, AND F. H. RAAB. *Solid State Engineering.* J. Wiley & Sons, 1980.

[2] E.W. PAPPENFUS, W.B. BRUENE, AND E.O. SCHOENIKE. *Single Sideband Principles and Circuits.* McGraw–Hill Book Company, New York, USA, 1964.

[3] F.N. SECHI. Linearised class–B transistor amplifiers. *IEEE Journal of Solid–State Circuits*, SC–11(2):264–270, 1976.

[4] A. USHIDA AND L.O. CHUA. Frequency–domain analysis of nonlinear circuits driven by multi–tone signals. *IEEE Transactions on Circuits and Systems*, CAS–31(9):766–778, September 1984.

[5] K.S. KUNDERT AND A.SANGIOVANNI-VINCENTELLI. Simulation of nonlinear circuits in the frequency domain. *IEEE Transactions of Computer–Aided Design*, CAD–5(4):521–535, October 1986.

[6] B. GILBERT. A monolithic microsystem for analog synthesis of trigonometric functions and their inverses. *IEEE Journal of Solid–State Circuits*, SC–17(6):1179–1191, December 1982.

[7] N. BLACHMAN. Detectors, bandpass nonlinearities, and their optimization: Inversion of the Chebyshev transform. *IEEE Transactions on Information Technology*, IT–17:398–404, July 1971.

[8] R. SEA AND A. VACROUX. Steady–state analysis of non–linear systems and multiple input describing functions. *Automatica*, 5:763–772, 1969.

[9] D.D. ŠILJAK. *Nonlinear systems.* J. Wiley & Sons, New York, USA, 1969.

[10] A. KAYE, D. GEORGE, AND M. ERIC. Analysis and compensation of bandpass nonlinearities for communications. *IEEE Transactions on Communications*, COM–20:965–972, October 1972.

[11] P. HETRAKUL AND D.P. TAYLOR. Compensators for bandpass nonlinearities in satellite communications. *IEEE Transactions on Aerospace and Electronic Systems*, 12(4):509–514, July 1976.

[12] W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY. *Numerical recipes in C – The art of scientific computing.* Cambridge University Press, second edition, 1992.

[13] S.M. PIZER. *Numerical computing and mathematical analysis.* Science Research Associates, 1975.

[14] C.F. GERALD AND P.O. WHEATLEY. *Applied numerical analysis.* Addison–Wesley Publishing Company, New York, USA, fifth edition, 1994.

[15] Y.L. KUO. Noise loading analysis of a memoryless nonlinearity characterized by a Taylor series of finite order. *IEEE Transactions on Instrumentation and Measurement*, IM–22(3):246–249, September 1973.

[16] R.S. LARKIN. Multiple signal intermodulation and stability considerations in the use of linear repeaters. In *IEEE 41st Vehicular Technology Conference*, pages 747–752, St Louis, Missouri, USA, May 1991.

[17] R.G. SEA. An algebraic formula for the amplitudes of intermodulation products involving an arbitrary number of frequencies. *IEEE Proceedings*, 56:1388–1389, August 1968.

[18] T. NARHI. Analysis of strongly nonlinear circuits with a frequency–domain method coupled with a consistent large–signal model. In *IEEE International Microwave Symposium Digest (MTT-S)*, volume 2, pages 633–636, Atlanta, Georgia, June 1993.

[19] A.D. BOOTH. *Numerical methods.* Butterworth & Co., London, UK, third edition, 1966.

[20] A. A. M. SALEH. Frequency–independent and frequency–dependent nonlinear models of TWT amplifiers. *IEEE Transactions on Communications*, COM–29(11):1715–1720, November 1981.

[21] W.B. RIBBENS. Spectral response of a nonlinear device. *Proceedings of the IRE*, 49(11):1700–1701, November 1961.

[22] P.M. BAKKEN. A new method for computing the properties of the output signal of nonlinear amplifiers. *IEEE Transactions on Communications*, COM–34(1):67–71, January 1986.

[23] C.M. CHIE. A modified Barrett–Lampard expansion and its application to bandpass nonlinearities with both AM–AM and AM–PM conversion. *IEEE Transactions on Communications*, COM–28(11):374–382, November 1980.

[24] O. SHIMBO. Effects of intermodulation, AM–PM conversion and additive noise in multicarrier TWT systems. *IEEE Proceedings*, 59:230–238, February 1971.

[25] R. BLUM AND M.C. JERUCHIM. Modeling nonlinear amplifiers for communication simulation. In *IEEE International Conference on Communications*, pages 1468–1472, Boston, USA, June 1989.

[26] M.T. ABUELMA'ATTI. Frequency–dependent nonlinear quadrature model of TWT amplifiers. *IEEE Transactions on Communications*, COM–29(11):982–986, August 1984.

[27] N. WIENER. *Nonlinear problems in random theory.* Technology Press, 1958.

[28] M.C. JERUCHIM, P. BALABAN, AND K.S. SHANMUGAN. *Simulation of Communications Systems*, pages 141–163. Plenum Press, New York, USA, 1992.

[29] C.L. LAW AND C.S. AITCHISON. Prediction of wide–band power performance of MESFET distributed amplifiers using the Volterra series representation. *IEEE Transactions on Microwave Theory and Techniques*, MTT–34(12):1308–1317, December 1986.

[30] S. NARAYANAN. Transistor distortion analysis using Volterra series representation. *Bell System Technical Journal*, 46:991–1024, May–June 1967.

[31] S.A. MAAS. *Nonlinear Microwave Circuits*. Artech House, 1988.

[32] B.J. LEON AND D.J. SCHAEFER. Volterra series and Picard iteration for nonlinear circuits and systems. *IEEE Transactions on Circuits and Systems*, CAS–25(9):789–793, September 1978.

[33] J.G. MCRORY AND R. JOHNSON. Volterra kernel estimation for mildly nonlinear amplifiers. *Electronics Letters*, 29(23):2007–2008, November 1993.

[34] J.J. BUSSGANG, L. EHRMAN, AND J.W. GRAHAM. Analysis and nonlinear systems with multiple inputs. *Proceedings of the IEEE*, 62(8):1088–1119, August 1974.

[35] M.B. STEER AND P.J. KHAN. An algebraic formula for the output of a system with large–signal, multifrequency excitation. *Proceedings of the IEEE*, 71(1):177–179, January 1983.

[36] G.W. RHYNE, M.B. STEER, AND B.D. BATES. Frequency–domain nonlinear circuit analysis using generalized power series. *IEEE Transactions on Microwave Theory and Techniques*, MTT–36(2):379–387, February 1988.

[37] M.B. STEER, P.J. KHAN, AND R.S. TUCKER. Relationship between Volterra series and generalised power series. *Proceedings of the IEEE*, 71(12):1453–1454, December 1983.

Chapter Three

Linear Amplification Techniques

This chapter introduces the classes of operation of RF amplifiers, and briefly explains their power efficiency and linearity characteristics. Techniques which can be used to improve either the efficiency or the linearity of the amplifier are described, to enable an informed choice of the methods to be further investigated for the required application.

3.1 Introduction

RF amplification is implemented using a number of techniques, and these are categorised depending upon the mode of operation of the active device. The mode of operation, termed the amplifier *class*, determines the efficiency and linearity characteristics of the amplifier.

All practical RF amplifiers exhibit a degree of nonlinearity, and often this causes unwanted frequency components to be present in their output signals. To reduce these distortion products it is necessary to reduce the apparent nonlinearity of the amplifier, either by altering its design, or by applying a linearisation technique, which modifies the signals at the input or output of the amplifier such that the composite linearity is improved. Alteration of the amplifier design can only provide a modest improvement in linearity, and so to attain maximum performance a linearisation scheme is required.

There are many techniques which can be used to linearise an amplifier, and these are usually categorised into one of three groups; feedback, complementary distortion, and feedforward techniques. Within these categories it is often possible to further subdivide into narrowband and broadband techniques. A number of these techniques will be discussed with regard to their suitability for the proposed application.

An alternative method to achieve linear transmission is the process of RF synthesis, in which the RF signal is directly generated. This process could theoretically provide excellent efficiency and linearity characteristics.

The power efficiency characteristics of a given amplifier class may be improved using various methods; unfortunately this is usually at the expense of linearity.

3.2 Definition of Terms

3.2.1 Power Efficiency

There are two terms of power efficiency used in this work; instantaneous and average efficiency.

Instantaneous efficiency is defined as the ratio of the instantaneous output power, P_o, to the instantaneous dc power supplied to the amplifier, P_{dc} [1].

$$\eta_{INST} = \frac{P_o}{P_{dc}} \tag{3.1}$$

Note that 'instantaneous power' is defined as the square of instantaneous envelope level.

Average efficiency is defined as the ratio of the average output power, $P_{o,AVG}$, to the average dc power supplied to the amplifier, $P_{dc,AVG}$.

$$\eta_{AVG} = \frac{P_{o,AVG}}{P_{dc,AVG}} \tag{3.2}$$

The average efficiency is therefore dependent upon the characteristics of the input signal, in addition to the instantaneous efficiency. This effect will be further investigated in Chapter 5.

[1] This definition of instantaneous efficiency is in fact the definition for instantaneous collector (or drain) efficiency [1].

3.2.2 Amplifier Back–off

An amplifier is said to be *backed–off* when it is operated below its saturated output level. The degree to which it is backed–off may be referenced to either the input or output. A linear amplifier will have identical input and output back–off levels; this will not be the case with nonlinear amplifiers. A back–off of 0dB implies that the amplifier is operating at its saturated output power.

3.2.3 Bandwidth of operation

The bandwidth over which an amplifier operates is typically described as being either wideband, or narrowband. Such terms are not well defined, and so a more robust definition is required[2].

Relative bandwidth is expressed as a percentage, and is defined as

$$B_{rel} = \frac{B}{f_0} \tag{3.3}$$

where B is the absolute bandwidth, and f_0 is the centre frequency of the band. This is an important parameter in the characterisation of amplifiers and auxiliary components.

Channel bandwidth is the absolute bandwidth which a single channel occupies in a multi–channel system. It will be shown that this determines the linearisation scheme which may be used.

For the purposes of this work, the definitions in Table 3.1 are used.

	Relative Bandwidth	Channel Bandwidth
Narrowband	$< 0.5\%$	< 30kHz
Mediumband	0.5–5.0%	30–300kHz
Wideband	$> 5.0\%$	> 300kHz

Table 3.1: Bandwidth definitions.

3.3 Amplification Techniques

3.3.1 Introduction

There are four traditional amplifier classes, categorised by the fraction of the input cycle during which the active device conducts. These are the familiar classes A, B, AB and C [2].

More recently, a plethora of additional amplifier classes have been described, which in some applications can offer increased power efficiency over the traditional classes.

3.3.2 Traditional Amplifier Classes

The general form of a traditional amplifier is shown in Figure 3.1; in this case a common–emitter stage is shown, although other configurations such as common–base may be preferred in some situations. The circuit consists of an active device (assumed to be a transistor), a

[2]The author acknowledges Adrian Mansell as the source of this definition.

bias network to control the mode of operation, an RF choke (RFC) to enable power to be taken from the dc supply without loss of RF power, and a tuned circuit. The tuned circuit is not a prerequisite to operation for the linear classes, but provides harmonic suppression. Not shown in the diagram are the input and output matching circuits required at RF to provide efficient power transfer from the source and to the load. It is assumed throughout this section that the amplifier is perfectly matched.

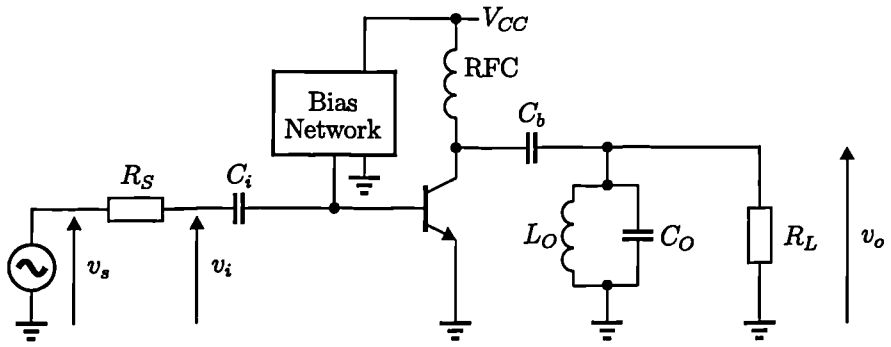


Figure 3.1: General form of a classical amplifier.

The operation of the four traditional classes will be illustrated with the use of their collector current waveforms during two RF cycles, shown in Figure 3.2. In the class AB and B cases the collector current shown is that through only one of the transistors.

Class A

The transistor is biased at a quiescent current, I_{CQ}, greater than, or equal to, the amplitude of the input signal, $\hat{I}_C$. Thus the transistor will conduct throughout the entire input cycle, giving a conduction angle of 360°.

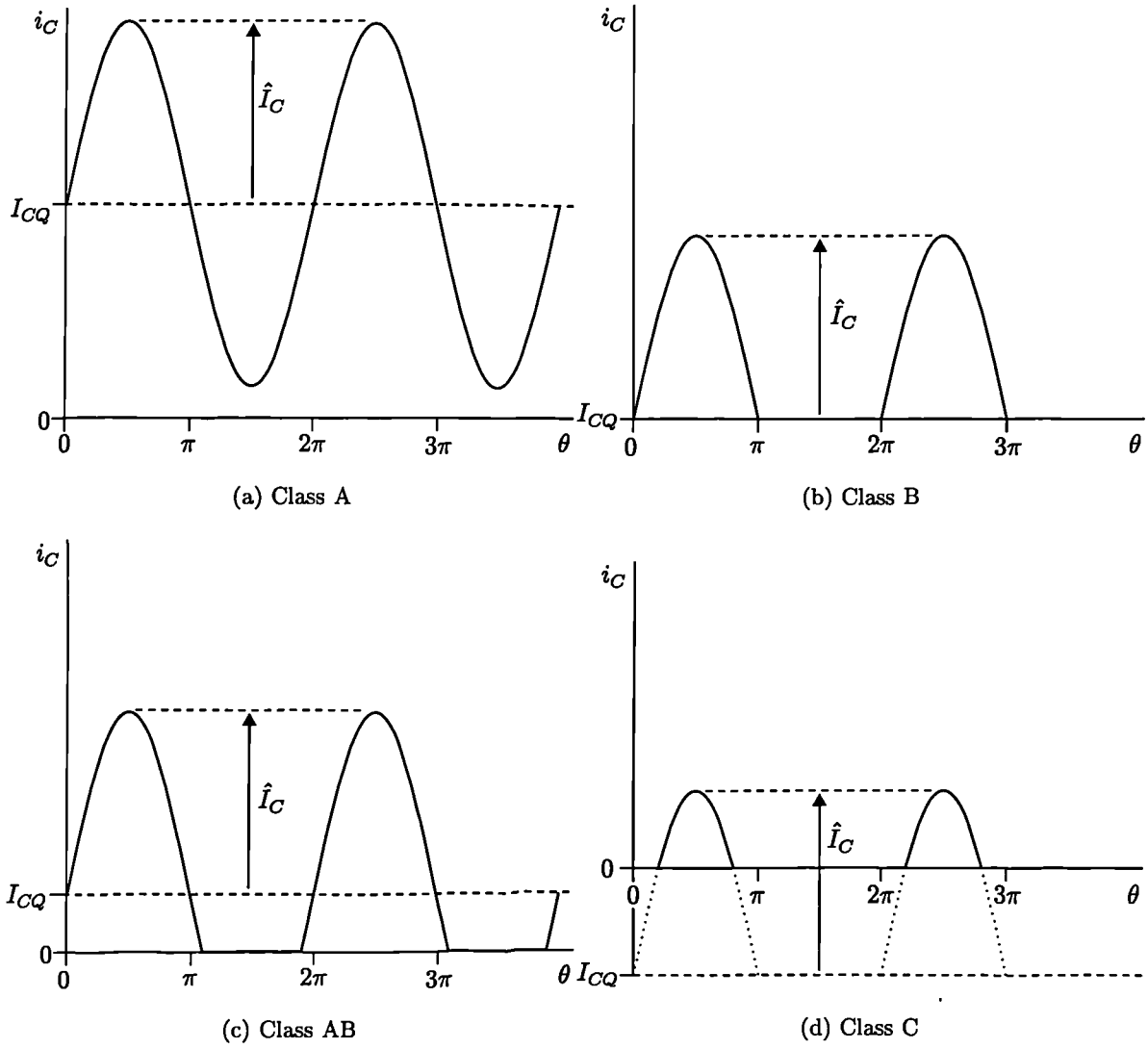
The maximum efficiency which can be attained occurs when the peak output current is equal to the bias current, and is theoretically 50%. In general, the actual efficiency will be significantly lower than this, often at RF in the range of 5–15%.

Although the efficiency is generally poor, this is the most linear of the traditional amplifier classes. A typical class A amplifier would produce IMPs of better than −30dBc in a two–tone test, rapidly improving as the output power of the amplifier is reduced; generally for each 1dB of amplifier backoff, the IMP level reduces by 2dBc (since the IMPs produced are predominantly of third–order).

Class B

In class B amplification the transistor conducts for half of the input cycle (*i.e.*, a conduction angle of 180°). Although it is possible to use only one transistor in a single–ended manner, the other half–cycle is usually generated by another transistor, in a push–pull arrangement. The efficiency is higher than can be attained with class A amplification, which is due to the lack of bias current (*i.e.*, $I_{CQ} = 0$), and reaches a theoretical maximum of 78.5%.

The linearity performance is lower than with class A amplifiers, due mainly to distortion which occurs as the transistor switches on after each zero–crossing (cross–over distortion). The transfer characteristic of the transistor is generally highly nonlinear at these very low levels, producing significant IMPs.



(a) Class A

(b) Class B

(c) Class AB

(d) Class C

Figure 3.2: Classical amplifier classes.

Class AB

Class AB amplification uses one or two transistors in the same manner as class B, but each transistor conducts for slightly more than half the cycle (*i.e.*, a conduction angle of greater than 180°), but much less than the entire cycle. This small amount of bias reduces the cross–over distortion which occurs with class B amplifiers, but reduces efficiency to some extent.

Class C

Class C amplifiers have a conduction angle significantly less than 180°, implemented by reverse–biasing the transistor, and have a maximum theoretical efficiency of 100%. The tuned–circuit is an integral part of the circuit operation, and forces the output voltage to be sinusoidal.

Solid–state class C amplifiers have a mode of operation which is significantly different to that of the original valve class C amplifiers, and has been termed mixed–mode class C by some

authors [1]. The circuit operation of mixed–mode class C amplifiers is highly complex, and varies considerably between amplifiers.

A typical class C amplifier may have an efficiency of 50–70%, and IMPs in a two–tone test of −15dBc. Since most of this distortion is due to the transistor switching on, rather than due to compression, the linearity performance degrades as the amplifier is backed–off.

Power Efficiency Characteristics

The theoretical instantaneous power efficiency of the traditional amplifier classes is shown in Figure 3.3. All of the classes except class B, have a range of possible efficiency characteristics. In class AB and C amplifiers this is due to the range of possible conduction angles which may be used. In the class A amplifier this is due to the use of a higher bias current than the peak output current, *i.e.*, $I_{CQ} > \hat{I}_C$.

As the amplifier output amplitude is reduced from its peak value (normalised to 1 in this case), the instantaneous power efficiency falls. Thus, when a non–constant envelope signal is amplified, the average efficiency is dependent not only upon the amplifier efficiency characteristic, but also upon the input signal statistics.

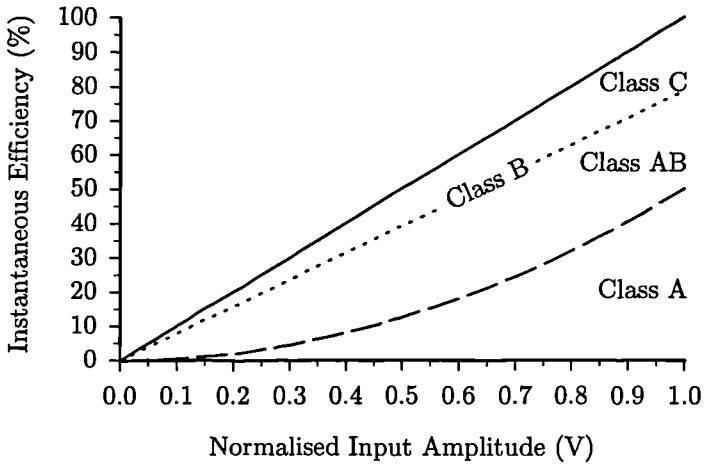


Figure 3.3: Ideal efficiency characteristics for traditional amplifier classes.

3.3.3 Linearity Characteristics

A comparison of the linearities of practical classes A and C amplifiers is shown in Figure 3.4. The class A amplifier (using a VNA025 MMIC) can be seen to become steadily more linear as the amplifier is backed–off from its PEP output, and can reach linearities such that the third–order IMPs are below −60dBc.

The class C amplifier (using the same amplifier as described in Chapter 2) has a more complex characteristic. At very low input levels (greater than 15dB back–off) the transistor is almost completely off, leading to a negligible output power. Between 15–12dB back–off the transistor is just starting to switch on, and so gain expansion is occurring, leading to reduced linearity as the input back–off is reduced. At less than 12dB back–off the transistor is switched on, and starts to enter saturation. In this region the linearity, in terms of third–order IMPs, increases with increasing drive level, although higher–order IMPs become more significant. As the amplifier becomes even more saturated the IMP level again rises with increasing output level.

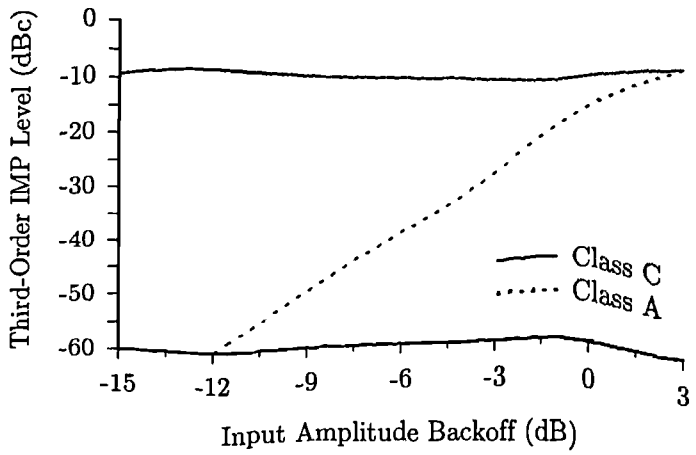


Figure 3.4: Third–order IMP level in Class A and Class C amplifiers.

3.3.4 Non–traditional Amplifier Classes

The efficiency of an amplifier may be increased by the use of techniques which reduce the average collector voltage–current product. This has led to the introduction of a number of alternative amplifier classes, some of which are briefly described in the following sections. It should be noted that the class designations have not yet been standardised, and thus are not universal.

Switching Amplifiers

One method to reduce the collector dissipation is to operate the transistor as a switch, rather than a current source (as it is used in traditional amplifier classes). This has led to the formulation of classes D, E and S [3–5], which perform this task in slightly different ways. The output of the amplifier is passed through a harmonic filter, to produce a first–zone output. All of these amplifier classes have a maximum theoretical efficiency of 100%, which is independent of output amplitude, *i.e.*, their ideal average efficiency will be 100%. No linearity can be ascribed to switching amplifiers due to their discontinuous transfer characteristics, and thus linearisation techniques may not be used. Linear transmission can still be attained using RF synthesis techniques.

Additional classes

There are other, non–switching, techniques which can be used to reduce collector dissipation compared to traditional amplifier classes. Although these amplifier classes can be more efficient than the traditional linear amplifier classes, the linearity performance is generally impaired.

In class F amplifiers [6] the load network is designed to resonate at the carrier frequency and one or more of its harmonics (generally the second or third harmonic). By correctly adjusting the amplitude and phase of these harmonics it is possible to flatten the collector voltage, thus improving efficiency.

Class G amplifiers [7] may be typically considered as consisting of two push–pull class B amplifiers operating from two supply voltages. Low–level signals are amplified by one pair of transistors, running from a low–voltage supply, with the second pair of transistors cut–off. At

some higher level signal, the first pair of transistors are cut–off and the signal is amplified by the second pair, fed by a higher–voltage supply. The average efficiency is, therefore, increased using this circuit topology.

Class H amplifiers [1] have a configuration similar to that of an amplifier with envelope–tracking (to be described in Section 3.3.5), but act on the instantaneous signal, rather than the envelope. The amplifier is similar to a class B amplifier, but the collector voltage is controlled, with the use of a high–efficiency switching amplifier, such that it is just higher than the output signal. Thus the collector dissipation is considerably reduced, improving the average efficiency.

3.3.5 Efficiency Enhancement Schemes

Envelope–tracking

Envelope–tracking [1] is similar to the RF synthesis technique Envelope–Elimination and Restoration, to be described in Section 3.8.2, and has the configuration shown in Figure 3.5. A highly–efficient switching amplifier is used to add an offset voltage to the supply voltage of a class B amplifier, such that the supply voltage is just large enough to allow linear amplification. The offset voltage is controlled by the amplitude of the input signal. Thus the power dissipation in the transistor is significantly reduced, improving the average efficiency of the amplifier compared to a class B amplifier.

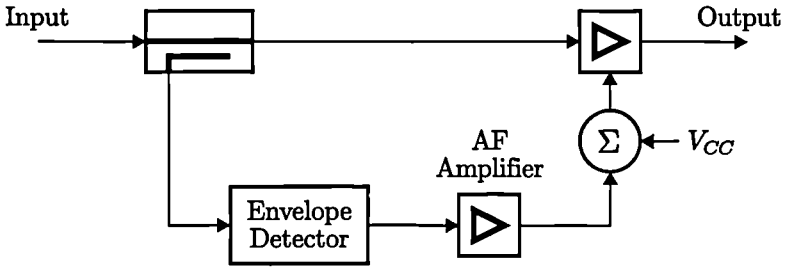


Figure 3.5: Configuration of an amplifier with envelope tracking.

Adaptive Bias

Adaptive bias [8], shown in Figure 3.6, is used to improve the average efficiency of class A amplifiers. The bias current is modulated by the envelope of the input signal, with low envelope levels causing a reduction in the bias current. Therefore, although the peak efficiency will be unchanged compared to a class A amplifier, as the amplifier is backed–off the efficiency will not be reduced by the same degree. It is likely, however, that the linearity will not be as high as with a class A amplifier.

Doherty

The Doherty technique [9, 10] uses a small number of linear amplifiers whose output is combined through an impedance–inverting coupler. A two–amplifier Doherty system is shown in Figure 3.7, which will be used to explain its operation. At low envelope levels PA_1 operates linearly, with PA_2 switched off. As the envelope level increases due to modulation, PA_1 will

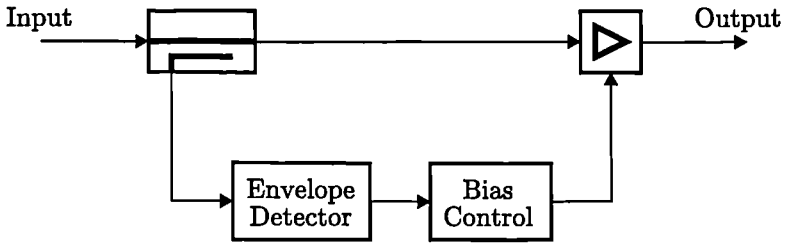


Figure 3.6: Configuration of an amplifier with adaptive bias.

eventually saturate (at some factor, α, of the maximum envelope), and PA_2 will operate linearly. The net effect is to produce a linear amplifier, with an improved average efficiency over conventional techniques.

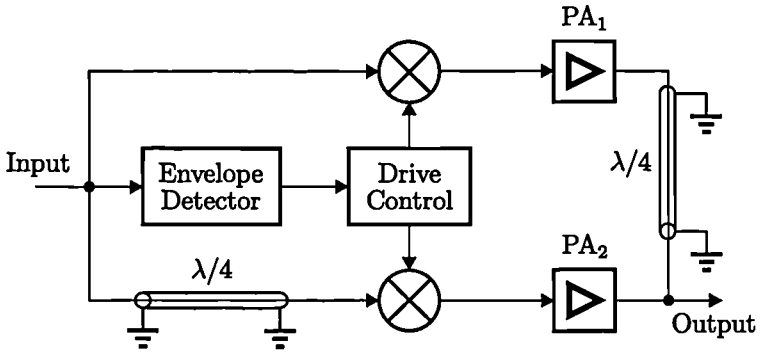


Figure 3.7: Configuration of a two–amplifier Doherty system.

The instantaneous efficiency of a Doherty system using two idealised class B amplifiers, with $\alpha = 0.5$, is shown in Figure 3.8. The instantaneous efficiency is significantly greater for the Doherty system than for the class B amplifier, and so the average efficiency when amplifying non–constant envelope signals is also much increased.

The impedance–inverting coupler is usually realised with a quarter–wave transmission line, and hence limits the operational bandwidth of the Doherty system.

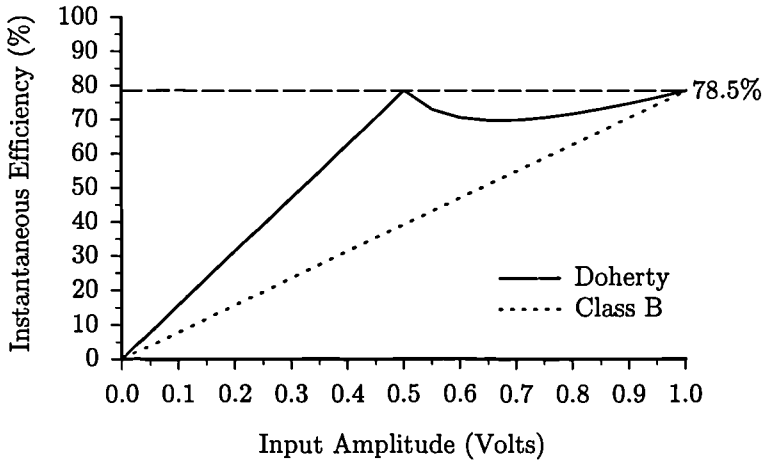


Figure 3.8: Instantaneous efficiency of a Doherty system using two amplifiers ($\alpha = 0.5$).

3.4 Amplifier Linearisation Techniques

To improve the linearity of the amplifier topologies described in the previous section, a linearisation technique may be used. Linearisation techniques fall into three broad categories; feedback, complementary distortion and feedforward methods. Feedback and complementary distortion techniques are both quasi–linear, *i.e.*, they modify the input signal to force the amplifier to produce the required linear output. This has consequences on their efficiency, which becomes dependent upon input amplitude. This effect will be further investigated in Chapter 5. Feedforward linearisation injects an additional signal into the output signal, in order to cancel the distortion.

3.5 Feedback Techniques

Feedback techniques utilise a sample of the output signal, which is passed through a feedback network and subtracted from the input signal. The feedback amplifier gain, G_f is given by

$$G_f = \frac{G}{1 + G\beta} \tag{3.4}$$

where G is the amplifier gain, and β is the feedback network gain. If the open–loop gain ($G\beta$) is large, the feedback amplifier gain will be almost exclusively determined by the feedback network, and is equal to the amount of attenuation in the feedback path ($1/\beta$). The reduction in IMPs is determined by the open–loop gain, assuming that the feedback network is linear, and that the amplifier does not saturate (in which case the open–loop gain reduces to almost zero).

Feedback is a continuous analogue process, which can compensate for changes caused by external effects, and so no adaptive control scheme is required, as is often necessary with other forms of linearisation. There are a number of possible system configurations.

3.5.1 RF Feedback

RF feedback, shown in Figure 3.9, is the simplest form of feedback, and all parts of the system are operating at the carrier frequency. The feedback network can be either passive or active.

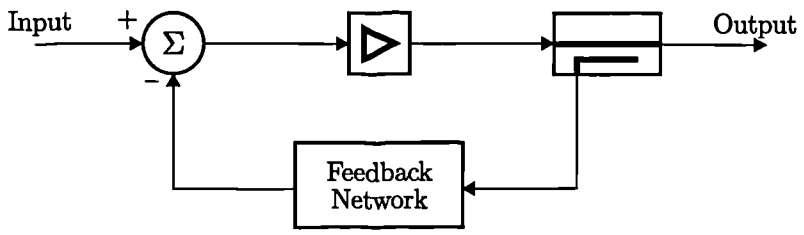


Figure 3.9: Configuration of an amplifier with RF feedback linearisation.

Passive RF Feedback

Passive RF feedback [11], uses a simple passive network as the feedback element, however its performance is limited by a number of factors. Firstly, as there is a finite delay around

the feedback loop, the bandwidth which can be linearised, without instability, is relatively narrow. To avoid positive feedback of out–of–band frequency components, a bandpass filter in the feedback path is usually required, increasing the delay and hence reducing the attainable bandwidth. The components in the amplifier are required to have a much larger bandwidth than the required operational bandwidth so that the feedback path delay is minimised; this increases the cost and complexity of the amplifier. Secondly, to achieve high levels of IMD cancellation requires the overall amplifier gain to be reduced by a similar extent. This is of particular importance at high frequency because of the reduced gain of the amplification devices. Therefore, it may be necessary to cascade a large number of stages to achieve the required gain and linearity.

Active RF Feedback

Active RF feedback [12–14], uses an active device, referred to as the auxiliary amplifier, in the feedback path. This auxiliary amplifier is designed to be relatively nonlinear and thus will introduce its own distortion to the sample of the output signal. The distorted signal is then added back into the input signal, and by controlling its amplitude and phase it is possible to reduce the main amplifier distortion.

The active RF feedback method has the advantage that the gain of the amplifier need not be reduced by the same extent as in passive RF feedback to achieve similar levels of linearity. This effect occurs because the auxiliary amplifier output contains more distortion than the main amplifier output and thus, when added back into the input, the distortion can be cancelled by a larger degree than the linear signal.

Using this technique, Ballesteros *et al* [14] constructed an amplifier which had a gain reduction of only 3dB, whilst achieving third–order IMD cancellation of 12dB. The distortion cancellation was found to be dependent upon the signal level, with only a small improvement at low and high signal levels.

3.5.2 Modulation Feedback

The most significant problem with RF feedback is that it is generally only possible to have a small amount of feedback, in order to avoid instability, resulting in only modest amounts of distortion reduction (of the order of 10 to 20dB). Modulation feedback reduces this problem by using the modulation as the fed–back signal which, in a narrowband system, is at a much lower frequency than the RF carrier. By making the feedback components' bandwidths much larger than the highest frequency component of the modulation, it is possible to apply a large amount of feedback whilst maintaining stability, and, therefore, produce a high degree of linearity. The modulation signal can be fed–back in either polar or cartesian form and so two methods of modulation feedback can be implemented.

Polar Loop Feedback

Polar loop feedback [15–17] feeds the modulation back in magnitude and phase components as shown in Figure 3.10. The difference between the magnitude of the input and output modulation is used to control the gain of the main amplifier, usually by modulating the amplifier supply voltage. The main amplifier input is driven from a VCO which is controlled by the phase difference between the input and the output modulation. This also has the effect of upconverting the modulation signal, and hence the technique forms a complete linear

transmitter, rather than merely a linear amplifier. The amplifier is very similar in form to the Envelope Elimination and Restoration (EE&R) amplifier technique explained in Section 3.8.2, and, therefore, polar loop feedback is often used to linearise an EE&R amplifier.

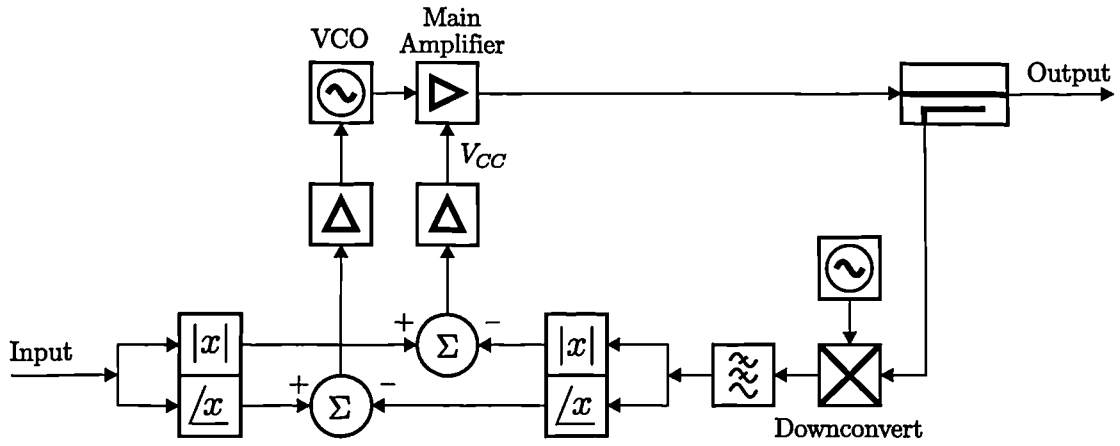


Figure 3.10: Configuration of an amplifier with polar loop feedback.

The linearity performance of an amplifier can be significantly improved over a narrow bandwidth by using the polar loop feedback technique. Petrovic [16] achieved intermodulation distortion products of below −55dBc in a two–tone test at a frequency of 168MHz with 1kHz tone separation. The main limitation of distortion cancellation was found to be due to imperfections in the polar resolver circuitry. The polar representation of the input signal leads to wider band signals than is the case with a cartesian representation, since discontinuities may result. The technique is best suited to modulation schemes with low envelope variation, to avoid noise problems in the envelope detection circuitry.

Envelope feedback [18–20] is a simplified form of polar loop feedback with only the envelope of the output signal fed–back; no phase information is returned. The difference between the input and output envelopes is used to control the gain of the amplifier such that the distortion is minimised. The performance of the technique is inferior to that of polar loop feedback, since only amplitude information is used in the feedback path; Smithers [20], achieved IMD below −40dBc in a two tone test.

Cartesian Loop Feedback

In contrast to the polar loop technique, cartesian loop feedback [21–26] resolves the modulation into I and Q vectors rather than magnitude and phase signals. A general configuration for the technique is shown in Figure 3.11. The input and output cartesian signals are subtracted to provide a loop error signal, which is then upconverted and applied to the amplifier input.

There are a number of inherent problems which must be overcome in order to achieve high linearity performance; loop stability, generation of accurate cartesian signals and maintenance of the quadrature of the local oscillators The loop stability can be improved by the use of a phase–shifter between the LO inputs of the quadrature up– and down–conversion paths, to increase the phase margin of the loop. Since the phase shift required is dependent on frequency, stability can only be maintained over a narrow bandwidth, and must be altered if the carrier frequency changes significantly, *e.g.*, due to a channel switch. A similar scheme can be used to facilitate maintenance of quadrature of the local oscillators [22]. A digital

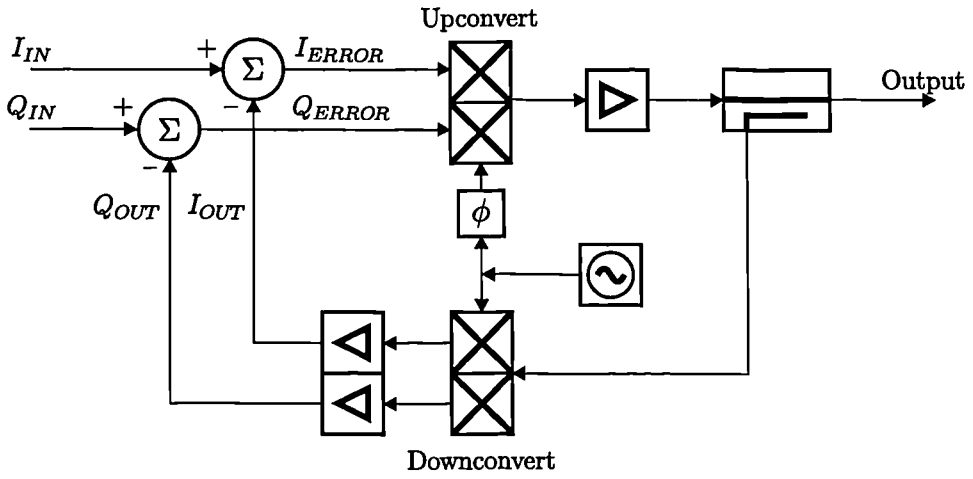


Figure 3.11: Configuration of an amplifier with cartesian loop feedback.

signal processor (DSP) is used to provide the cartesian modulation input, and this can be accurately controlled and maintained by using a calibration procedure.

The technique has been used to provide excellent linearity over a narrow bandwidth at various carrier frequencies. Wilkinson *et al* [21] achieved IMD cancellation of 45dB at 900MHz and 30dB at 1.7GHz using a two–tone input with 5kHz spacing. It is possible to use the technique over a wider band, but the loop gain has to be reduced to allow stable operation, and therefore the linearity performance is degraded compared with narrowband operation. Johansson *et al* [23,24] have achieved IMD reduction, in a two–tone test at 900MHz, of 29dB with 20kHz tone spacing, and 20dB with 1MHz tone spacing. Cartesian loop feedback amplifiers have also been constructed for HF operation [25], and VHF operation [26] with similar levels of performance.

3.5.3 Distortion Feedback

Distortion feedback [27,28], shown in Figure 3.12, combines a sample of the output signal with a delayed version of the input signal so that, with appropriate gain and phase adjustments, the input signal is cancelled. This leaves only distortion products which are then subtracted from the input signal. This results in cancellation of the distortion, without reducing the linear gain of the amplifier. However, only modest improvements in linearity have thus far been published; Gajda *et al* [27] achieved a reduction of third–order IMPs of 6dB over a 10MHz bandwidth centred at 300MHz.

The method used to obtain the distortion products is the same as the first loop in a feedforward system (Section 3.7), and is not self–adapting. Thus, an additional adaption mechanism may be required in practice, further reducing the attractiveness of this technique.

3.6 Complementary Distortion

Complementary distortion linearisation techniques utilise a nonlinear circuit either before or after the amplifier, known as predistortion and postdistortion respectively. The transfer characteristic of this nonlinear circuit is chosen such that the distortion which it introduces cancels that produced by the amplifier.

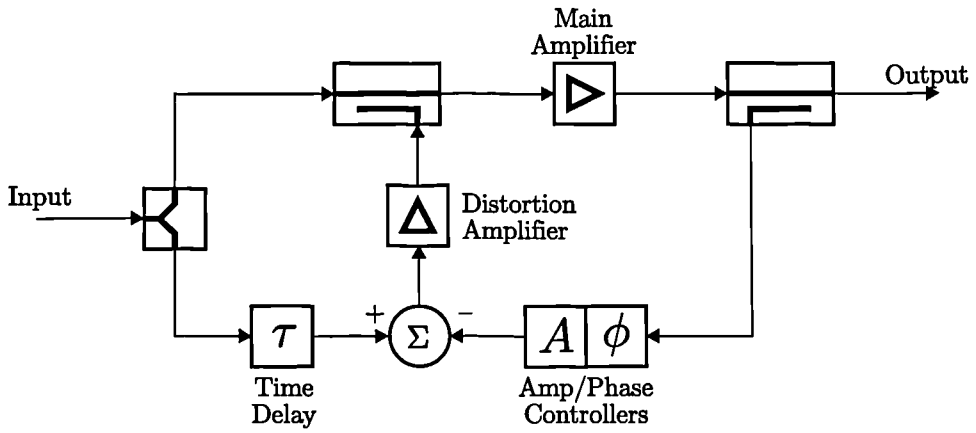


Figure 3.12: Configuration of an amplifier with distortion feedback.

3.6.1 Predistortion Linearisation

The general block diagram of an amplifier with predistortion is shown in Figure 3.13. The feedback techniques described in the previous section can be envisaged as a form of predistortion, but with continuous adaption via the feedback path. However, predistortion is an open–loop, and thus unconditionally stable, technique which does not inherently adapt to changes in the system characteristics, caused by external effects such as temperature changes or channel switching. Therefore, to achieve the optimal linearity performance, it is necessary to use a method to adapt the predistorter circuitry. Whilst this is normally implemented when baseband predistortion is used, it is less common with RF or IF predistortion due to the increased complexity, and more modest linearity improvement possible with these techniques.

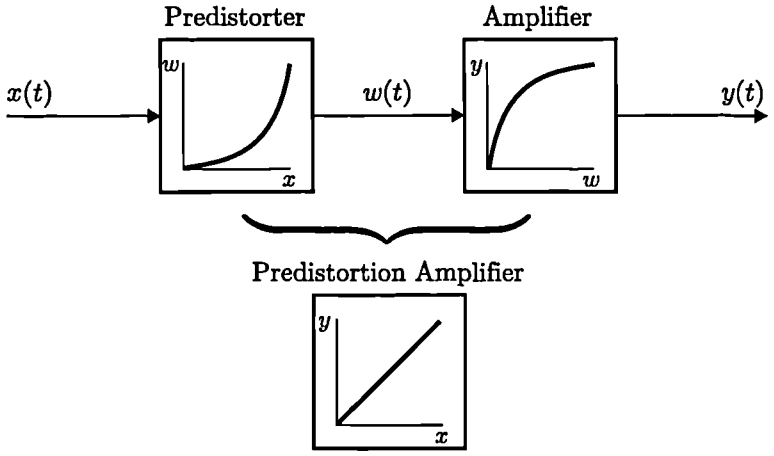


Figure 3.13: Configuration of an amplifier with predistortion.

3.6.2 Adaptive Baseband Predistortion

Adaptive baseband predistortion uses a DSP to calculate the coefficients of the predistorter required to produce a linear output. There are three basic forms; mapping–based, gain–based and analogue, the configuration of which is summarised in Figure 3.14.

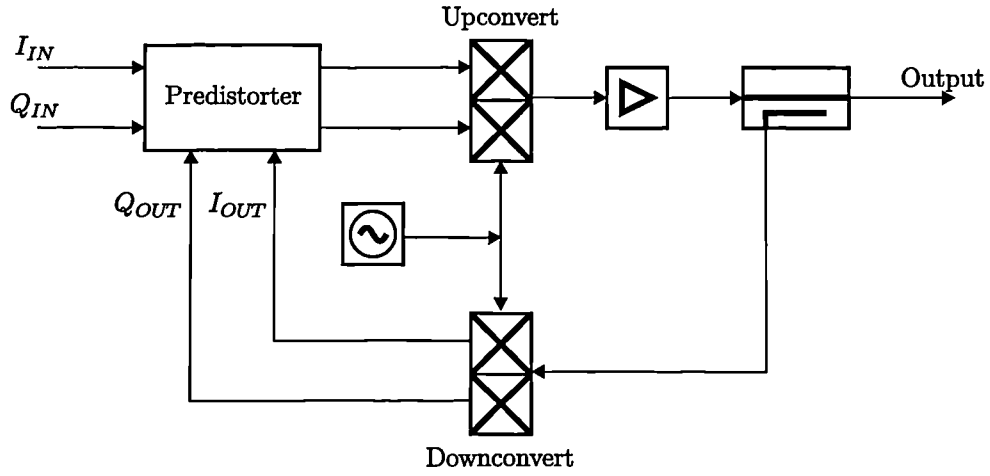


Figure 3.14: Configuration of an amplifier with adaptive baseband predistortion.

Mapping–Based Predistorter

The mapping–based predistorter method [29] uses a two–dimensional look–up table which maps any point on the complex plane to any other point. The input signal in cartesian form is converted by the DSP into a predistorted signal by direct mapping; this is then upconverted and amplified. A measure of the output signal is downconverted and used by the DSP to adapt the look–up table to produce the minimum error between the input and output signals in the cartesian plane, and therefore the minimum distortion. The direct mapping process allows compensation of any order of memoryless distortion, including errors produced by the modulation process [30]. The linearity of the method is potentially excellent, with out–of–band emissions reduced to below −60dBc (simulated).

The method does, however, have a number of disadvantages. The memory requirement is huge because every point on the complex plane has to be mapped; for 10bit accuracy, a look–up table of 20Mbits is required. The convergence of the predistorter is slow (typically tens of seconds) because every region in the cartesian plane must be accessed before convergence is complete, and the predistorter must reconverge completely on power–up, or at each channel change. One method of reducing the memory requirement and the convergence time is to use a constellation–based predistorter, which is a simplified version of the mapping–based technique. The look–up table is reduced from storing the entire complex plane to mapping the constellation points into a predistorted version. This also reduces the convergence time as less values in the look–up table have to be updated. However, this method can obviously only be used with digital modulation and, unlike the mapping–based predistorter, is not modulation–independent.

Gain–Based Predistorter

This method [31–33] is effectively a one–dimensional mapping predistorter, and uses the envelope level of the signal to produce the required complex output, by storing a table of level–dependent complex gain values. (This is an acceptable method of reducing AM/AM and AM/PM conversion because, by definition, these are dependent upon the signal envelope only). Interpolation can be used to calculate intermediate gain values. The memory requirement of this method is considerably smaller than that required by the mapping–based method; typically only approximately 64 words are required to produce similar performance.

The convergence time is significantly shorter (of the order of milliseconds – dependent upon the channel bandwidth), and reconvergence time is negligible.

The computation required by this method is far more complex than with the mapping predistorter, and accurate control of the cartesian modulator is required, as the method cannot correct errors due to the modulation process.

Baseband Analogue Predistorter

This method [34] uses an analogue technique to predistort the input signal whose characteristic is controlled by a DSP. The predistorter characteristic is not altered in real time so the convergence time is relatively slow. The output of the amplifier is downconverted to baseband and the envelope is used directly by the DSP to calculate the new predistorter characteristic. No practical results have thus far been published; simulated results published [34] show only a 10dB reduction in IMD. However, the complexity is significantly reduced compared with the other adaptive baseband predistortion methods.

It should be noted that the use of a DSP for processing the envelope level means that the method, in its published form, is inherently narrowband.

3.6.3 RF and IF Predistortion

RF predistortion refers to the predistortion network being implemented at the carrier frequency, whereas IF predistortion is implemented at some intermediate frequency and then upconverted to the carrier frequency. The predistortion element is usually in analogue form to achieve the required bandwidths; however, digital wideband predistortion is possible.

Digital IF Predistortion

The use of digital predistortion networks to achieve wideband predistortion has been proposed [35]. The predistortion is performed on the digitised IF signal, and the method can be made adaptive if required. The results presented show a reduction in the third–order IMPs of 25dB, with a maximum bandwidth of 45MHz. The performance of this method is limited by the speed and accuracy of the converters, and the power requirements of such fast devices will limit the overall efficiency of the system.

RF and IF Analogue Predistortion

These are the preferred methods for achieving wideband predistortion of amplifiers, and use analogue components in the predistorter elements. The two types of predistorter (RF and IF) are equally popular, although it has been suggested [36] that IF predistortion may give superior performance since it enables a reduction in the nonlinear effects of the upconverter to be obtained. However, an IF predistorter is required to operate over a larger relative bandwidth than an RF predistorter and, therefore, may have inferior broadband performance.

The characteristics of the predistorter are generally non–adaptive, unlike in baseband predistortion, and therefore the performance does tend to degrade with environmental effects. Some adaptive RF and IF predistortion techniques have been proposed [37–39], although these tend to be relatively unsophisticated, and only allow slow adaption.

The performance of RF and IF predistortion tends to be limited, with a typical improvement in IMP level in a two–tone test of 10 to 20dB [40–42]. At present RF and IF predistortion is generally used in satellite systems to predistort TWTs, and television systems to predistort class AB amplifiers. Both of these types of amplifier have characteristics which become more linear as the output power is reduced. Thus, it is possible to improve the linearity, but reduce efficiency and power handling capabilities, by backing–off the amplifier. This is not the case with class C amplifiers, and thus the accuracy of the predistortion network is required to be considerably higher than with more linear amplifiers. This appears to have thus far prohibited the use of analogue predistortion of class C amplifiers.

3.6.4 Postdistortion Linearisation

The postdistortion linearisation technique [43] is similar to predistortion, but the distortion characteristic is implemented after the amplifier, rather than before it. The general block diagram form is shown in Figure 3.15. The technique is less satisfactory than predistortion because of the required large signal handling capability of the postdistortion block and the fact that any loss through it causes significant degradation of efficiency.

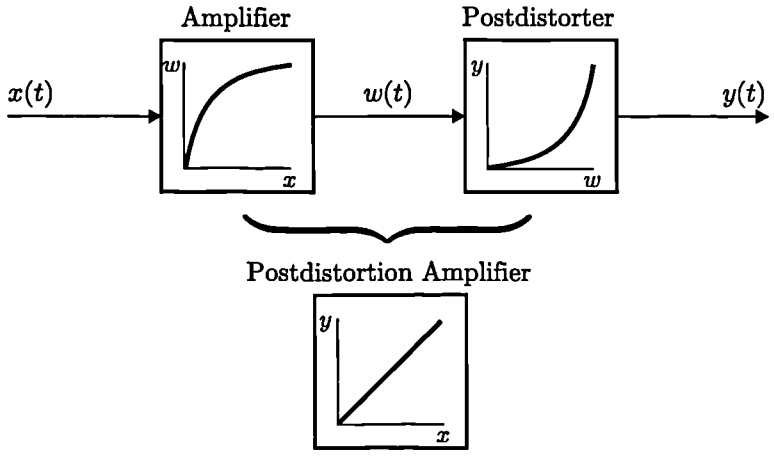


Figure 3.15: Configuration of an amplifier with postdistortion.

3.7 Feedforward Linearisation

The feedforward linearisation technique [44–48] is described with reference to a two–tone input as shown in Figure 3.16. The input signal is split to provide two paths; a high power main path (top) and a lower power error path (bottom). The main path signal is amplified by the main amplifier (A1), which introduces distortion. A portion of the main path signal, sampled via directional coupler C1, is subtracted from a delayed version of the input signal. This produces an error signal which is predominantly made up of distortion products. The error signal is amplified by the error amplifier (A2) and recombined via directional coupler C2, in antiphase with a delayed version of the main amplifier output. This has the effect of cancelling the distortion at the output. The gain and phase controllers equalise the signals at the two subtraction elements so as to attain maximum cancellation.

Further distortion cancellation may be achieved with additional feedforward loops, by considering the feedforward amplifier as the main amplifier in another feedforward system.

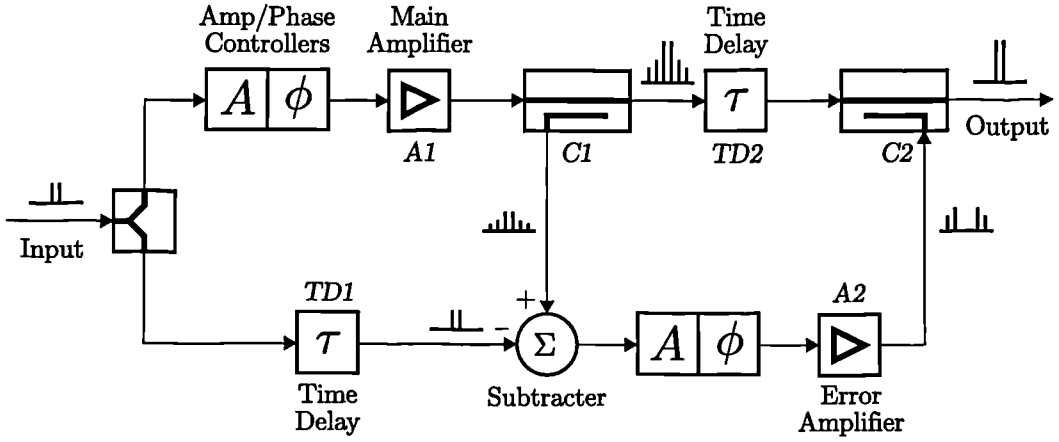


Figure 3.16: Configuration of a feedforward amplifier with a two–tone input.

The technique is open–loop and thus unconditionally stable, but requires adaption of the amplitude and phase of both paths to achieve optimal performance, and a number of adaption schemes have been proposed [49–51]. The linearity performance which can be achieved is excellent and can be maintained over a wide bandwidth. Kenington *et al* [52,53] have achieved a reduction in intermodulation distortion in excess of 60dB over a 30MHz band centered at 900MHz (using two feedforward loops).

3.8 RF Synthesis Techniques

RF synthesis techniques have been proposed to address the problem of low distortion RF amplification, by directly creating a magnified representation of the input signal at the required output power. Thus, they can provide linear transmission with the potential to achieve very high power efficiencies.

3.8.1 LINC

The LINC (LInear amplification using Nonlinear Components) technique, [54–57], also known as 'outphasing' [58], theoretically offers the potential of amplification with 100% efficiency over all power levels. It relies on the fact that a bandpass signal with amplitude and phase modulation, $S(t)$, is equivalent to the sum of two phase–modulated, constant–envelope signals, $S_1(t)$ and $S_2(t)$. Therefore, the signals $S_1(t)$ and $S_2(t)$ can be amplified by nonlinear, power efficient amplifiers and summed at a high level to produce a linear, high power version of the input signal. The basic LINC configuration is shown in Figure 3.17.

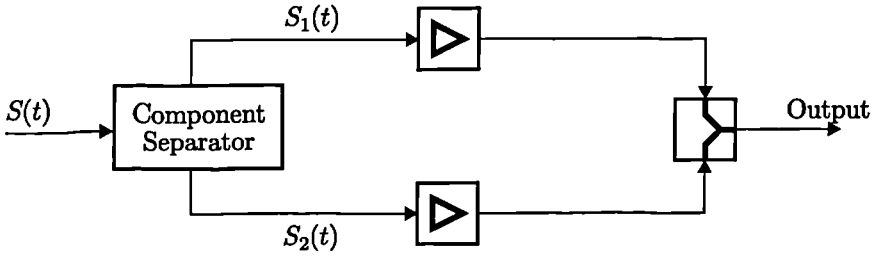


Figure 3.17: Configuration of a LINC amplifier.

ENGINEERING

However, the efficiency and linearity of the LINC amplifier are dependent upon a number of factors. Firstly, the nonlinear amplifiers used will not be 100% efficient. At VHF and UHF it is possible to use switching amplifiers, whose efficiency approaches 100%; however, at higher frequencies this is not possible and so class C amplifiers would be used. These typically have an efficiency of 60–70% at VHF, reducing to less than 40% at frequencies greater than 1GHz. Secondly, the high–level combination of the amplifier outputs will cause a significant amount of power loss, and therefore reduce efficiency.

The linearity of the amplifier is heavily dependent upon the accuracy of the signal separation process, particularly the generation of an accurate $\cos^{-1}[\ ]$ function, and on the amplitude and phase match of the two signals at the input to the final subtracter. These factors have thus far limited the performance of the LINC amplifier using analogue processing; Cox *et al* [59] attained intermodulation products of −40dBc over a 1MHz bandwidth. The use of digital signal processing, which can perform these functions accurately has been applied to good effect [60,61]. However, this restricts the LINC technique to narrowband systems.

Recently, a modified version of the LINC technique, known as CALLUM (Combined Analogue Locked Loop Universal Modulator), has been proposed [62], which allows a good level of performance whilst using simple and cheap circuitry.

The split signals, $S_1(t)$ and $S_2(t)$, occupy a wider band than the original modulation, dependent upon the envelope characteristics of the signal. The LINC signal bandwidth is often theoretically infinite, but is usually defined as the bandwidth outside of which signals are 60dB lower than the largest frequency component. Using this definition, the ratio of the LINC bandwidth to the modulation bandwidth is approximately 6 for $(\pi/4)$–DQPSK, 32 for QPSK, and 220 for a two–tone test[3]. This reduces the applicability of the LINC technique to wideband systems.

3.8.2 Envelope Elimination and Restoration

Envelope Elimination and Restoration (EE&R) [63] utilises two highly efficient amplifiers. The RF input signal is split into two parts, one with purely phase modulation (*i.e.*, RF) and the other with purely amplitude modulation (*i.e.*, the AF envelope). These are then amplified by RF and AF power amplifiers respectively, as shown in Figure 3.18. The RF signal is constant–envelope, and therefore the amplifier used can be highly nonlinear and thus highly efficient. The AF amplifier is used to modulate the supply voltage to the RF amplifier to produce a linear output. The AF amplifier can also be highly efficient as it only needs to amplify a low frequency signal, thus a PWM switching amplifier with output filter is often used.

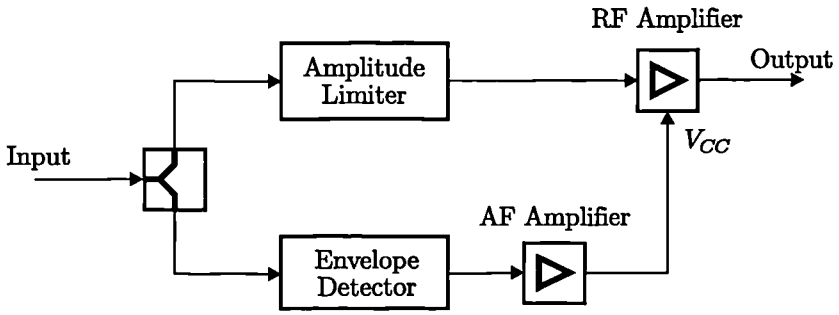


Figure 3.18: Configuration of an envelope elimination and restoration amplifier.

[3]The author gratefully acknowledges Adrian Mansell for his provision of these figures.

The overall power efficiency is heavily dependent upon the efficiency of the RF amplifier. At frequencies in and below the UHF range it is possible to use switching amplifiers, which have excellent efficiency characteristics. However, as with the LINC technique, these amplifiers cannot be used at higher frequencies, and hence class C amplifiers would be used, with a consequent reduction in efficiency.

Using an EE&R amplifier, Koch *et al* [64] achieved an average efficiency exceeding 50% whilst maintaining third–order IMPs below −30dBc in a two–tone test. Linearisation of the amplifier can be achieved with the use of modulation feedback, and the circuit topology is particularly suited for the use of polar loop feedback [65].

The AF signal will generally occupy a wider band than the original modulation, limiting the technique's wideband performance.

3.9 Summary

This chapter has described the linearity and power efficiency characteristics of the various classes of amplifier. In general, the more efficient an amplifier, the less linear it will be, and vice versa. Therefore, to attain highly efficient linear amplification some form of linearisation scheme, or an RF synthesis technique would be required.

To this end, a number of the most significant forms of linearisation techniques for RF amplifiers, and two RF synthesis techniques have been introduced. This project requires the design of an amplifier suitable for satellite and cellular applications, and therefore demands high frequency, wideband operation with good linearity and efficiency characteristics.

RF feedback has two major limitations; it cannot provide significant linearisation without excessive loss in gain, and the operational bandwidth is limited by stability considerations. The linearity can be increased by using modulation feedback techniques, but these are inherently narrowband. The loss of linear gain can be reduced by using active or distortion feedback methods, but these provide only modest linearity improvement. Therefore, whilst feedback techniques do not require additional circuitry for adaption, they are only conditionally stable, and hence cannot easily provide good linearity over a wide bandwidth.

The adaptive baseband predistortion techniques can potentially provide excellent improvements in linearity, but their use of a DSP to process the input signal limits them to narrowband systems.

The postdistortion method has limited efficiency performance due to its use of analogue circuitry after the power amplifier, which will introduce loss.

The LINC and EE&R amplifier techniques can provide excellent efficiency, with reasonable linearity, at frequencies below UHF. Above this, it is not possible to use RF switching amplifiers and so the efficiency will be reduced. The high power recombination of the signals in a LINC amplifier will also degrade efficiency. Using the EE&R amplifier in a wideband system requires that the envelope amplifier be wideband also, and therefore an AF PWM switching amplifier cannot be used, and the alternatives will reduce the power efficiency. The linearity obtained using these techniques is dependent upon the accuracy of the signal splitting and recombination. Using analogue techniques provides only modest linearity which can be improved using a DSP to split the signal. This, however, limits the technique to narrowband applications. Both of the techniques suffer from the fact that the signals used occupy a wider band than the original modulation.

Only three techniques remain, which offer the potential of wideband linear operation at

high frequencies; IF and RF predistortion and feedforward linearisation, and these will be investigated further.

Due to the added complexity of the efficiency enhancement schemes for RF amplifiers, these will not be further investigated. However, the linearisation techniques chosen for further study do not preclude their use in a future system.

REFERENCES

[1] H. L. KRAUSS, C. W. BOSTIAN, AND F. H. RAAB. *Solid State Engineering.* J. Wiley & Sons, 1980.

[2] T. HA. *Solid-State Amplifier Design.* J. Wiley & Sons, New York, USA, 1969.

[3] W.J. CHUDOBIAK AND D.F. PAGE. Frequency and power limitations of class-D transistor amplifiers. *IEEE Journal of Solid-State Circuits*, SC-4(1):25-37, February 1969.

[4] N.O. SOKAL AND A.D. SOKAL. Class-E — A new class of high-efficiency tuned single-ended switching power amplifiers. *IEEE Journal of Solid-State Circuits*, SC-10(3):168-176, June 1975.

[5] M. KAZIMIERCZUK. Exact analysis of class E tuned power amplifier with only one inductor and one capacitor in the load network. *IEEE Journal of Solid-State Circuits*, SC-18(2):214-221, April 1983.

[6] D.M. SNIDER. A theoretical analysis and experimental confirmation of the optimally loaded and overdriven RF power amplifier. *IEEE Transactions on Electron Devices*, ED-14(12):851-857, December 1967.

[7] F.H. RAAB. Average efficiency of class-G power amplifiers. *IEEE Transactions on Consumer Electronics*, CE-32(2):145-150, May 1986.

[8] A.A.M. SALEH AND D.C. COX. Improving the power-added efficiency of FET amplifiers operating with varying-envelope signals. *IEEE Transactions on Microwave Theory and Techniques*, MTT-31(1):51-56, January 1983.

[9] W.H. DOHERTY. A new high efficiency power amplifier for modulated waves. *Proceedings of the IRE*, 24(9):1163-1182, September 1936.

[10] F.H. RAAB. Efficiency of Doherty RF power-amplifier systems. *IEEE Transactions on Broadcasting*, BC-33:77-83, September 1987.

[11] A.F. MITCHELL. A 135 MHz feedback amplifier. In *IEE Colloquium on Broadband High Frequency Amplifiers - Practice and Theory*, pages 2.1-2.6, London, UK, November 1976.

[12] F. PEREZ, E. BALLESTEROS, AND J. PEREZ. Linearisation of microwave power amplifiers using active feedback networks. *Electronics Letters*, 21(1):9-10, January 1985.

[13] J.C. PEDRO AND J. PEREZ. An MMIC linearized amplifier using active feedback. In *IEEE International Microwave Symposium Digest (MTT-S)*, volume 1, pages 95–98, Atlanta, Georgia, June 1993.

[14] E. BALLESTEROS, F. PEREZ, AND J. PEREZ. Analysis and design of microwave linearized amplifiers using active feedback. *IEEE Transactions on Microwave Theory and Techniques*, 36(3):499–504, March 1988.

[15] V. PETROVIC AND W. GOSLING. Polar–loop transmitter. *Electronics Letters*, 7:286–287, May 1979.

[16] V. PETROVIC AND C.N. SMITH. The design of VHF SSB polar–loop transmitters. In *IEE Conf. on Communications, Equipment and Systems*, volume CP 209, pages 148–155, Birmingham, UK, April 1982.

[17] V. PETROVIC. A high efficiency VHF SSB transmitter. In *Proceedings of the Conference on Civil Land Mobile Radio*, pages 47–55, London, UK, November 1975.

[18] T. ARTHANAYAKE AND H.B. WOOD. Linear amplification using envelope feedback. *Electronics Letters*, 7(7):145–146, April 1971.

[19] W.B. BRUENE. Distortion reducing means for single–sideband transmitters. *Proceedings of the IRE*, 44:1760–1765, December 1956.

[20] C.R. SMITHERS. Further reflections on envelope feedback for linearising bipolar power amplifiers at HF and VHF. In *IEE Int. Conf. on Mobile Radio Systems and Techniques*, volume CP 238, pages 270–280, York, UK, September 1984.

[21] R.J. WILKINSON, J. MACLEOD, M.A. BEACH, AND A. BATEMAN. Linear transmitter design for MSAT terminals. In *2nd International Mobile Satellite Conference*, pages 297–301, Ottawa, Ontario, Canada, June 1990.

[22] A. BATEMAN AND D.M. HAINES. Direct conversion transceiver design for compact low–cost portable mobile radio terminals. In *IEEE 39th Vehicular Technology Conference*, pages 1/57–1/62, San Francisco, California, May 1989.

[23] M. JOHANSSON AND T. MATTSSON. Transmitter linearization using cartesian feedback for linear TDMA modulation. In *IEEE 41st Vehicular Technology Conference*, pages 439–444, St. Louis, Missouri, May 1991.

[24] M. JOHANSSON AND T. MATTSSON. Linearised high–efficiency power amplifier for PCN. *Electronics Letters*, 27(9):762–764, April 1991.

[25] V. PETROVIC. Reduction of spurious emission from radio transmitters by means of modulation feedback. In *IEE 2nd Int. Conf. on Radio Spectrum Conservation Techniques*, volume CP 224, pages 44–49, Birmingham, UK, September 1983.

[26] V. PETROVIC. VHF SSB transmitter employing cartesian feedback. In *IEE Conference on Telecommunications, Radio and Information Technology*, volume CP 235, pages 161–165, Birmingham, UK, May 1984.

[27] G.B. GAJDA AND R.J.P. DOUVILLE. A linearization system using RF feedback. In *IEEE International Electrical and Electronics Conference*, pages 30–33, Toronto, Canada, 1983.

[28] R. GÓMEZ AND H. JARDÓN. Highly linear amplifier for high gain applications. *Electronics Letters*, 32(2):81–82, January 1996.

[29] Y. NAGATA. Linear amplification technique for digital mobile communications. In *IEEE 39th Vehicular Technology Conference*, pages 159–164, San Francisco, California, May 1989.

[30] M. FAULKNER, T. MATTSSON, AND W. YATES. Adaptive linearisation using pre–distortion. In *IEEE 40th Vehicular Technology Conference*, pages 35–40, Orlando, Florida, May 1990.

[31] M. FAULKNER, M. JOHANSSON, AND W. YATES. Error sensitivity of power amplifiers using pre–distortion. In *IEEE 41st Vehicular Technology Conference*, pages 451–456, St. Louis, Missouri, May 1991.

[32] J.K. CAVERS. Amplifier linearization using a digital predistorter with fast adaption and low memory requirements. *IEEE Transactions on Vehicular Technology*, 39(4):374–382, November 1990.

[33] A.R. MANSELL AND A. BATEMAN. Practical implementation issues for adaptive predistortion transmitter linearisation. In *IEE Colloquium on Linear RF Amplifiers and Transmitters*, pages 5/1–5/7, April 1994.

[34] S.P. STAPLETON AND J.K. CAVERS. A new technique for adaption of linearizing predistorters. In *IEEE 41st Vehicular Technology Conference*, pages 753–758, St Louis, Missouri, USA, May 1991.

[35] M. HORN AND A. EGGER. Design and performance of microwave predistortion networks using digital circuits. In *14th European Microwave Conference*, pages 549–554, Belgium, September 1984.

[36] R.P. HECKEN, R.C. HEIDT, AND D.E. SANFORD. Predistortion for the travelling–wave–tube amplifier. *Bell System Technical Journal*, 62(10):3447–3465, December 1983.

[37] HIDEO ASHIDA, YOSHIAKI SUZUKI, ISAMU UMINO, AND NORIO TOZAWA. C–Band 100 Watt GaAs FET amplifier for digital radio. In *IEEE International Conference on Communications*, pages 1/23–1/27, Seattle, Washington, June 1987.

[38] J. NAMIKI. An automatically controlled predistorter for multilevel quadrature amplitude modulation. *IEEE Transactions on Communications*, COM–31(5):707–712, May 1983.

[39] T. NOJIMA AND T. KONNO. Cuber predistorter linearizer for relay equipment in 800 MHz band land mobile telephone system. *IEEE Transactions on Vehicular Technology*, VT–34(4):169–177, November 1985.

[40] M. KUMAR, J.C. WHARTENBY, AND H.J. WOLKSTEIN. Predistortion linearizer using GaAs dual–gate MESFET for TWTA and SSPA use in satellite transponders. *IEEE Transactions on Microwave Theory and Techniques*, MTT–33(12):1479–1488, December 1985.

[41] S. KUMAR. Power amplifier linearization using MMICs. *Microwave Journal*, pages 96–104, April 1992.

[42] G.M. BLAIR, J.J. DALY, AND J.F. MOSS. Design and characterisation of a GaAs FET power amplifier for a 64 QAM digital radio transmitter. In *IEEE International Conference on Communications*, volume 1, pages 1.4.1–1.4.6, Seattle, Washington, June 1987.

[43] A. PROCHAZKA, P. LANCASTER, AND R. NEUMANN. Amplifier linearization by complementary pre– or post–distortion. *IEEE Transactions on Cable Television*, CATV–1(1):31–39, October 1976.

[44] T.J. BENNETT AND R.F. CLEMENTS. Feedforward – an alternative approach to amplifier linearisation. *The Radio and Electronic Engineer*, 44(5):257–262, May 1974.

[45] J.P. DIXON. A solid–state amplifier with feedforward correction for linear SSB applications. In *IEEE International Conference on Communications*, volume 1, pages 728–732, Toronto, Canada, June 1986.

[46] A. JAVED, P.A. GOUD, AND B.A. SYRETT. Analysis of a microwave feedforward amplifier using Volterra series representation. *IEEE Transactions on Communications*, COM–25(3):355–360, March 1977.

[47] P.D. LUBELL AND W.B. DENNISTON. Linearizing amplifiers for multi–signal use. *Microwaves*, pages 46–50, April 1974.

[48] R.G. MEYER, R. ESCHENBACH, AND W.M. EDGERLEY. A wide–band feedforward amplifier. *IEEE Journal of Solid–State Circuits*, SC–9(6):422–428, December 1974.

[49] R.M. BLUMENKRANZ AND P. ENG. A microprocessor monitored feed–forward super trunk CATV amplifier. *IEEE Transactions on Cable Television*, CATV–5(3):105–116, July 1980.

[50] S. NARAHASHI AND T. NOJIMA. Extremely low–distortion multi–carrier amplifier — self–adjusting feed–forward (SAFF) amplifier —. In *IEEE International Conference on Communications*, volume 3, pages 1485–1490, Denver, June 1991.

[51] J. POWELL AND T. HA. *UK Patent GB 2 238 197 A*, May 1991.

[52] P.B. KENINGTON, R.J. WILKINSON, AND J.D. MARVILL. A multi–carrier amplifier for future mobile communications systems. In *IEE International Conference on Mobile Radio and Personal Communications*, pages 151–156, Coventry, UK, December 1991.

[53] P.B. KENINGTON, R.J. WILKINSON, AND J.D. MARVILL. Broadband linear amplifier design for a PCN base–station. In *IEEE 41st Vehicular Technology Conference*, pages 155–160, St. Louis, Missouri, USA, May 1991.

[54] D.C. COX. Linear amplification with nonlinear components. *IEEE Transactions on Communications*, COM–22:1942–1945, December 1974.

[55] D.C. COX AND R.P. LECK. Component signal separation and recombination for linear amplification with nonlinear components. *IEEE Transactions on Communications*, pages 1281–1287, November 1975.

[56] S. TOMISATO, K. CHIBA, AND K. MUROTA. Phase error free LINC transmitter. *Electronics Letters*, 25:576–577, April 1989.

[57] F. CASEDEVALL AND J.J. OLMOS. On the behaviour of a LINC transmitter. In *IEEE 40th Vehicular Technology Conference*, pages 29–34, Orlando, Florida, USA, May 1990.

[58] F.H. RAAB. Efficiency of outphasing RF power–amplifier systems. *IEEE Transactions on Communications*, COM–33:1094–1099, October 1985.

[59] D.C. COX AND R.P. LECK. A VHF implementation of a LINC amplifier. *IEEE Transactions on Communications*, COM–24:1018–1023, September 1976.

[60] S.A. HETZEL, A. BATEMAN, AND J.P. MCGEEHAN. LINC transmitter. *Electronics Letters*, 27:844–845, May 1991.

[61] S.A. HETZEL, A. BATEMAN, AND J.P. MCGEEHAN. A LINC transmitter. In *IEEE 41st Vehicular Technology Conference*, pages 133–137, St Louis, Missouri, USA, May 1991.

[62] A. BATEMAN. The combined analogue locked loop universal modulator (CALLUM). In *IEEE 42nd Vehicular Technology Conference*, pages 759–763, Denver, Colorado, May 1992.

[63] L.R. KAHN. Single–sideband transmission by envelope elimination and restoration. In *Proceedings of the IRE*, pages 803–806, July 1952.

[64] M.J. KOCH AND R.E. FISHER. A high efficiency 835 MHz linear power amplifier for digital cellular telephony. In *IEEE 39th Vehicular Technology Conference*, pages 17–18, San Francisco, California, May 1989.

[65] F.H. RAAB. Envelope elimination and restoration system requirements. In *Green Mountain Radio Research, GMRR TP87-4*, pages 499–512.

Chapter Four

Feedforward Linearisation

Feedforward linearisation is a wideband technique, and as such has the potential to be used in the required application. This chapter includes new work which can be used to predict the linearity and efficiency performance of a non–ideal feedforward amplifier. Practical results of a feedforward amplifier using class C main and error amplifiers are also presented. Finally, methods which could improve the performance of the technique are examined.

4.1 Introduction

The feedforward linearisation technique can improve the linearity of an amplifier over a wide bandwidth. A high–level error signal is added to the output signal in order to cancel the distortion. The cancellation process is one of phasor addition, and hence to achieve good cancellation, careful gain and phase matching between the two signals is required. The high–level injected signal does not add any useful output power, and so its generation reduces the overall efficiency.

This chapter theoretically analyses the linearity and power efficiency performance of a non–ideal feedforward amplifier. Practical verification of the analyses are presented, using a VHF feedforward amplifier with class C main and error amplifiers. Methods are described for improving the efficiency of the feedforward amplifier. Finally, a technique combining feedforward linearisation and analogue predistortion is presented, which offers the potential of increased linearity and efficiency compared with feedforward linearisation alone.

4.2 Linearity Aspects of a Feedforward Amplifier

4.2.1 Introduction

The block diagram of a feedforward amplifier is shown in Figure 4.1. The cancellation of distortion by the feedforward process is achieved by adding identical distortion to that produced by the amplifier in anti–phase to the output signal. The accuracy with which this can be performed will give an ultimate limit to the cancellation possible, and hence to the linearity of the final output. It is the precise matching in amplitude and phase through the main and error loops which is crucial, and this must be maintained beyond the entire operating bandwidth, since intermodulation products will spread outside the band.

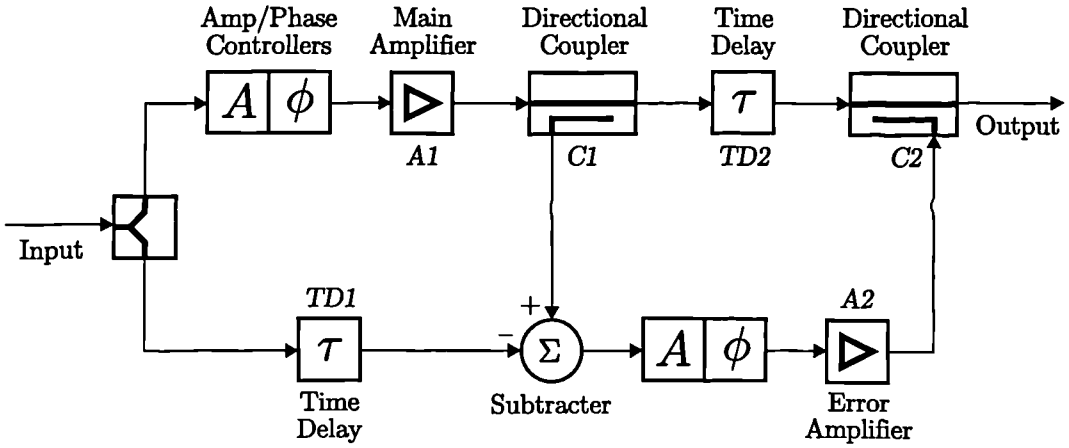


Figure 4.1: Configuration of a feedforward amplifier.

There are a number of factors which affect the amplitude and phase matching in a practical system. The frequency–dependent nature of the components in both paths (most notably the amplifiers and amplitude and phase controllers) cause alteration of the amplitude and phase across the frequency band. Secondly, the nonlinearity of the error amplifier causes changes in the amplitude and phase of the error signal, dependent upon the envelope level.

This section analyses the linearity of a feedforward amplifier, by determining the cancellation

of two signals with different time delays, phases and amplitudes. Practical verification of the analysis is presented.

4.2.2 Typical Characteristics

The linearity of a feedforward amplifier is theoretically analysed in Appendix B (Section B.1). The cancellation which can be achieved is shown to be dependent upon the amplitude and phase differences between the two paths. The amplitude error is denoted by δA, and phase difference between the paths is given by $\theta + \delta\theta$, where θ is the gross phase shift, and $\delta\theta$ takes account of any further phase error. If the two paths are not exactly time matched, the delay mismatch ($\Delta\tau$) reduces cancellation across the band. The cancellation which can be attained at an angular frequency ω is given by

$$C_{dB} = -10\log_{10}\left\{1 + (1+\delta A)^2 + 2(1+\delta A)\cos\left(\omega\Delta\tau - \theta - \delta\theta\right)\right\} \tag{4.1}$$

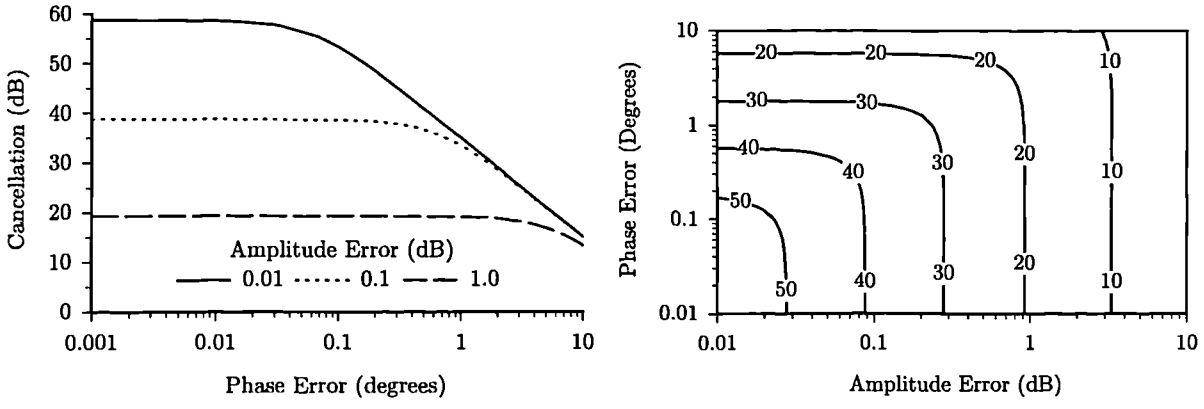
and the phase difference between the input and the cancelled output, $\Delta\gamma$, is given by

$$\Delta\gamma = \tan^{-1}\left(\frac{(1+\delta A)\sin\left(\omega\Delta\tau - \theta - \delta\theta\right)}{1 + (1+\delta A)\cos\left(\omega\Delta\tau - \theta - \delta\theta\right)}\right) \tag{4.2}$$

There are a number of cases which must be considered to illustrate the analysis fully.

Cancellation with no delay mismatch

In a practical feedforward system the delay mismatch is usually incorrectly assumed to be zero, by the use of compensating delay lines, and the gross phase shift assumed to be 180°. The cancellation which can be attained is thus limited by the amplitude and phase errors between the two paths (δA and $\delta\theta$ respectively) and is shown in Figure 4.2(a), and in contour form in Figure 4.2(b). The matching required is stringent; to achieve 30dB of distortion cancellation requires an amplitude match of better than 0.3dB, and phase matching to within 2°.



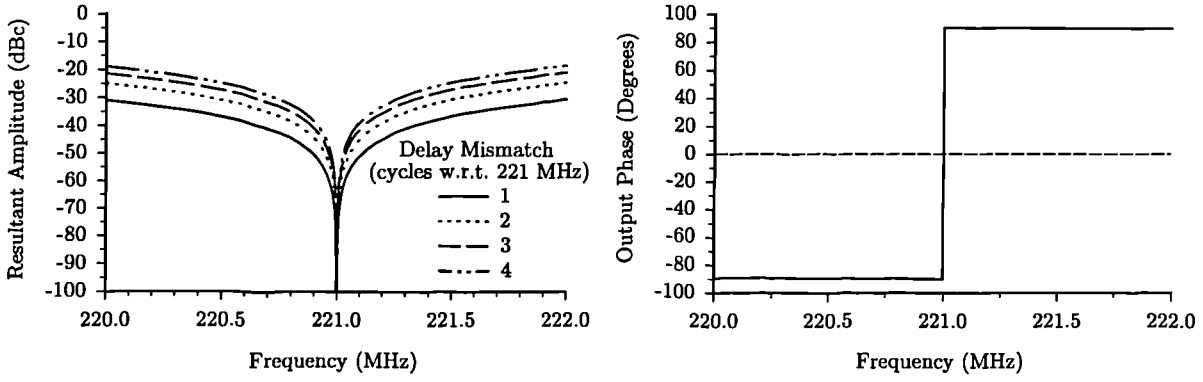
(a) Standard Form (b) Contour Form

Figure 4.2: Cancellation between signals applied to the inputs of a combiner with amplitude and phase errors.

Cancellation with delay mismatch, but no amplitude or phase errors

If there is some delay mismatch between the two cancellation paths, it is not possible to attain the required 180° phase difference for ideal cancellation, at all frequencies. However, it is possible to have phase matching at one frequency (ω_0), usually arranged to be the centre of the band, where perfect cancellation will occur (*i.e.*, $\omega_0 \Delta\tau = n\pi$). At frequencies other than this the cancellation will reduce.

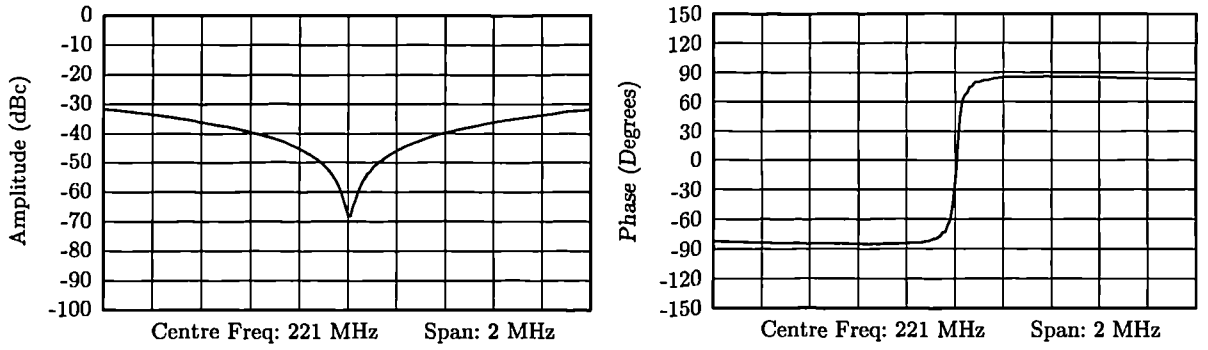
The resultant output, when two signals are combined, across a 2MHz band at a centre frequency of 221MHz is shown in Figure 4.3(a), with delay mismatches of one to four cycles with respect to the centre frequency, with no amplitude or phase errors. The cancellation is at a maximum at the centre frequency and decreases towards the extremes of the band and as the number of cycles of delay mismatch is increased. However, greater than 30dB of cancellation is possible across the entire band with one cycle of delay mismatch. The phase difference between the input and output is shown in Figure 4.3(b). From the extremes of the band to the centre frequency the phase difference tends to ±90°, whilst at the centre frequency there is a step in phase of 180°.



(a) Resultant output

(b) Phase shift (one cycle of delay mismatch)

Figure 4.3: Effect of delay mismatch on resultant output and phase shift across 2MHz band, at a centre frequency of 221MHz, with delay mismatch as a parameter.



(a) Resultant output

(b) Phase shift

Figure 4.4: Practical resultant output and phase shift across 2MHz band with centre frequency of 221MHz, and one cycle of delay mismatch.

This analysis has been experimentally verified, and the results for the 220–222MHz band are shown in Figure 4.4 for one cycle of delay mismatch. It can be seen that the results are in

close agreement with the theoretical results shown in Figure 4.3. The imperfect cancellation at the centre of the band is due to a small residual amplitude error.

The cancellation across the band is dependent upon the relative bandwidth, as shown in Figure 4.5 (with no amplitude or phase error). If cancellation, rather than reinforcement occurs (*i.e.*, positive cancellation) the cancellation decreases almost linearly with the logarithm of the relative bandwidth, and doubling the delay mismatch reduces the cancellation by approximately 6dB.

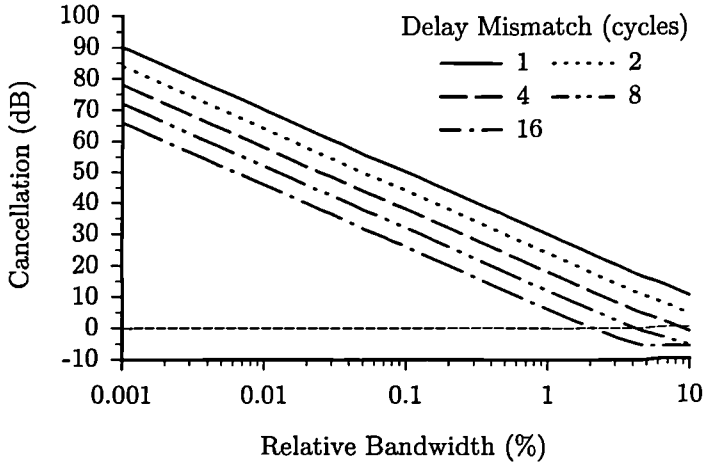


Figure 4.5: Cancellation versus relative bandwidth, with no amplitude or phase errors.

Cancellation with delay mismatch, and additional amplitude and phase errors

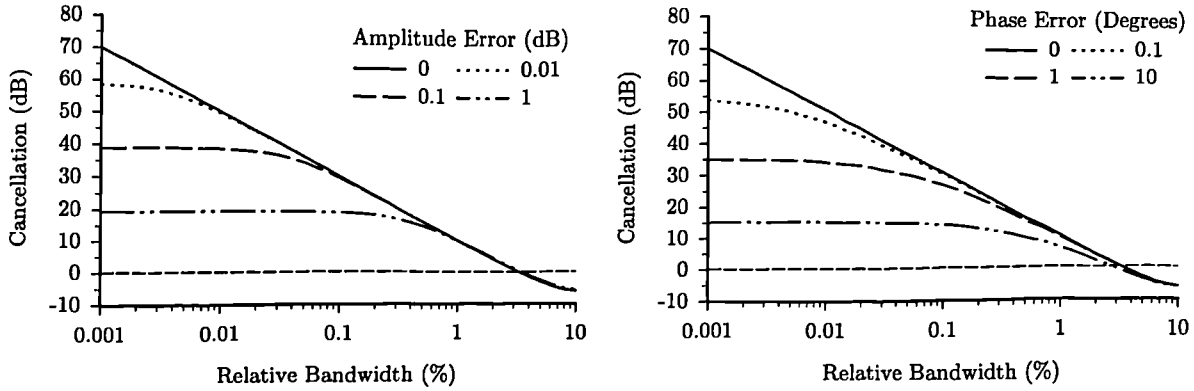
Figure 4.6 shows the effect which amplitude and phase errors have on the cancellation of the signals with some delay mismatch (10 cycles in this example). The additional errors will obviously reduce the amount of cancellation which can be attained. However, two distinct regions can be identified. Reducing the relative bandwidth from 10% improves the cancellation in a similar manner to that shown in Figure 4.5. Thus, the cancellation is determined largely by the amount of delay mismatch. However, as the relative bandwidth is further reduced, the cancellation becomes almost constant; in this region the amplitude error (or phase error) determines the cancellation. Therefore, for small relative bandwidths, the effects of delay mismatch on the cancellation may be negligible compared to those of the amplitude and phase errors. This is important in narrowband systems, since the delays may be reduced or eliminated without affecting the linearity, whilst (as will be shown in Section 4.3) significantly improving the power efficiency.

Cancellation with delay mismatch of less than $\lambda/2$

In standard feedforward systems, the delays are usually matched to within a half–wavelength ($\lambda/2$), and cancellation is achieved by adjusting the gross phase shift from 180°. The combination of this small delay mismatch and gross phase shift can achieve perfect cancellation at a single frequency, ω_0, in which case

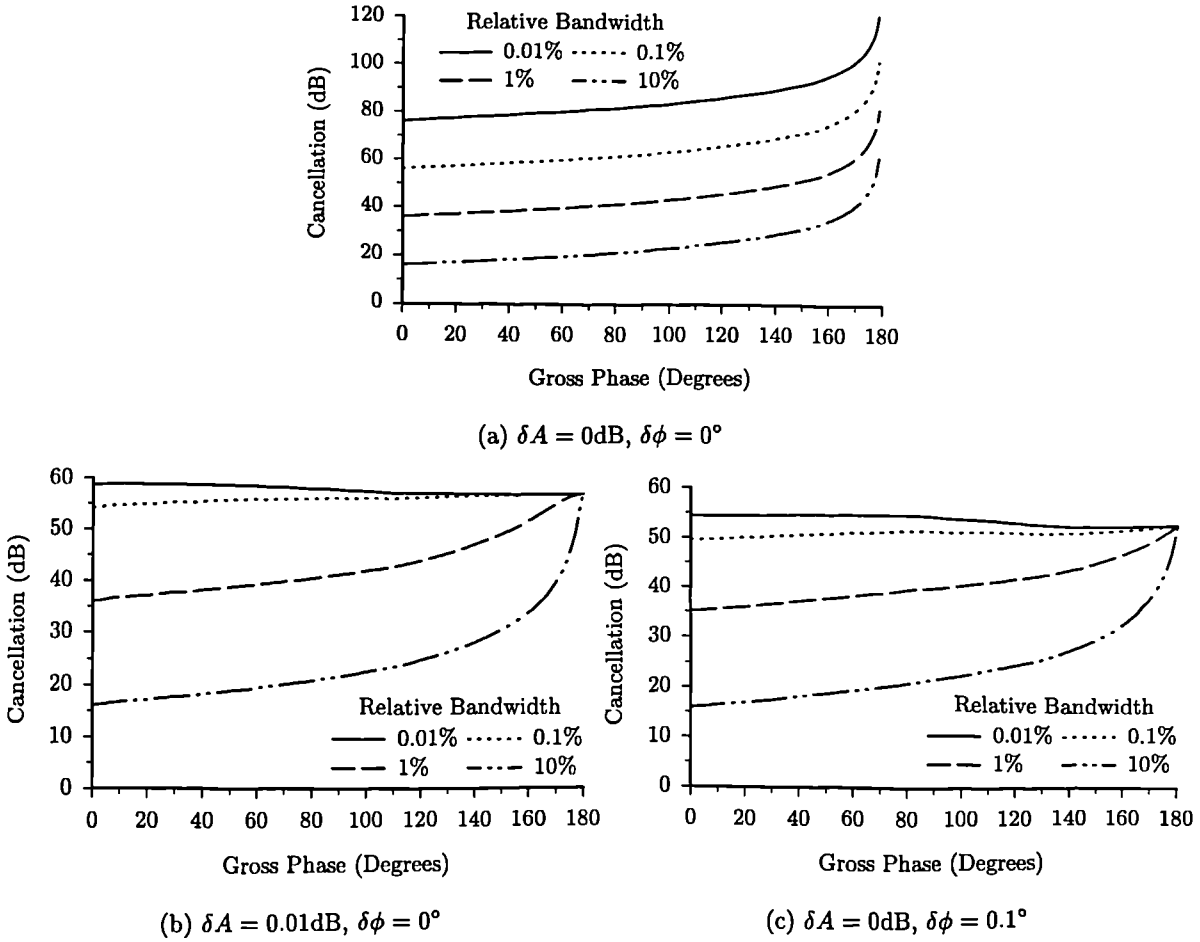
$$\omega_0 \Delta\tau - \theta = n\pi \tag{4.3}$$

The cancellation utilising a combination of gross phase shift and delay, with no additional amplitude or phase errors, is shown in Figure 4.7(a), and with some small additional amplitude



(a) Variable δA, $\delta\phi = 0°$ (b) $\delta A = 0$dB, Variable $\delta\phi$

Figure 4.6: Cancellation across relative bands, with 10 cycles of delay mismatch, and additional amplitude or phase errors.



(a) $\delta A = 0$dB, $\delta\phi = 0°$

(b) $\delta A = 0.01$dB, $\delta\phi = 0°$ (c) $\delta A = 0$dB, $\delta\phi = 0.1°$

Figure 4.7: Cancellation across relative bandwidths, with combination of phase and delay to achieve a 180° phase difference at the centre of the band.

and phase errors in Figures 4.7(b) and 4.7(c) respectively. The gross phase extends from 0 to 180°, requiring the delay mismatch to extend from $\lambda/2$ to 0, to compensate.

Without amplitude and phase errors the cancellation reduces rapidly if the gross phase shift,

θ, is not equal to 180° (*i.e.*, the delay mismatch is not zero). However, any amplitude or phase error reduces the degradation in cancellation if $\theta \neq 180°$. In a practical system there will always be some small amplitude and phase errors, hence the latter case is more representative. In practical narrowband systems the reduction in cancellation caused by a gross phase shift of other than 180° is negligible, and so the delays are only required to be relatively closely matched. This is not the case in practical wideband systems, since the degradation in cancellation can be significant. It is therefore important to match the delays as closely as possible in wideband feedforward systems.

4.2.3 Adaptive Control Schemes

The previous section has demonstrated how accurately the amplitude and phase balance between the cancelling signals in a feedforward system must be. Due to the open–loop nature of the feedforward technique, it cannot automatically compensate for any changes in amplitude or phase which would reduce linearity performance. For this reason, some means of adaptively controlling the amplitude and phase of the paths is generally required. In a dynamic sense, the speed of adaption may be required to be relatively slow, to compensate for the effects of temperature, or very rapid, to allow multiple channels to be switched on and off, or hopped in frequency, for example. To this end, a number of adaption techniques have been suggested, the basic principles of which will be briefly described. These schemes have no effect on the ultimate linearity of the feedforward system, but potentially allow the optimum linearity to be maintained at all times.

Pilot–tone adaption

This simple adaption scheme [1,2] uses a single frequency tone which is injected into the main signal path, and recovered at the feedforward amplifier output. As the signal is not present in the error (reference) path it should be cancelled in the same way as the main amplifier distortion. The amplitudes and phases are adjusted until the magnitude of the pilot–tone is minimised, and thus the distortion is minimised.

Whilst being simple, there are a number of disadvantages with the technique. Minimisation of the pilot–tone minimises the distortion around the pilot–tone frequency, but no account is taken of distortion at other parts of the band; thus the scheme is narrowband. Additional tones could be used throughout the band of operation, but this has obvious implications in complexity. The pilot–tone will occupy one of the channels which could be used for transmitting useful information, and so system capacity is reduced. In practice the technique does not adapt quickly, making its use in channel switching/hopping applications restrictive.

Signal correlation adaption

This method [3, 4] generally uses the signals already present throughout the feedforward amplifier to adapt the amplitude and phase controllers, and thus system capacity is unaffected. Correlation between signals is used to generate a complex error signal which enables the average cancellation of the distortion to be maximised.

This method allows transparent and rapid adaption of the feedforward amplifier. However, direct correlation, generating a dc error signal, is susceptible to dc offsets throughout the control circuitry (particularly the dc offset of the mixer used to correlate the signals). This disadvantage may be eliminated by downconverting the signals using slightly offset LO fre-

quencies [5, 6]. Correlation of these downconverted signals produces an error signal at a low frequency (typically a few kHz), which is then minimised, without being affected by dc offsets.

4.2.4 Reduction of Component Frequency–Dependence

The frequency–dependent nature of an amplifier and accompanying components degrades the level of distortion cancellation which can be attained with feedforward linearisation. It has been shown, however, [7] that the feedforward process reduces the frequency–dependent variation across the operating bandwidth. Thus, the use of an additional feedforward loop, termed a *zeroth loop*, has been suggested [6], prior to the standard feedforward technique. This would reduce the frequency–dependent variation of the amplitude and phase across the band, and improve distortion cancellation, at the cost of increased complexity.

4.3 The Efficiency of a Feedforward Amplifier

4.3.1 Introduction

The efficiency of a feedforward amplifier is inherently lower than that of the main amplifier, since the error amplifier is consuming power without producing useful output power (the output power is in fact reduced since the distortion is cancelled). The overall feedforward efficiency is not only affected by the efficiencies of the individual amplifiers, but is also highly dependent upon the error injection coupling factor, and the insertion loss of the time delay elements in the main signal path.

The effect of the error injection coupling factor has previously been examined [8]. The choice of coupling factor affects both the main and error path losses, and thus there is an optimal value which trades–off these losses to attain maximum overall efficiency. High values of coupling factor (*e.g.*, 20dB), produce low main path loss, but large error path loss, and thus significant power is required from the error amplifier, reducing overall efficiency. Low values of coupling factor (*e.g.*, 3dB), cause high main path loss, but small error path loss. Whilst this requires lower error amplifier power than the former case, the main amplifier power requirement is increased. It will be shown that there is an optimum coupling factor which compromises these two effects.

The finite insertion loss of the time delay elements in the main path result in significant power loss, and therefore reduced efficiency. This section extends the theory given in [8] to include the effects of both the error injection coupling factor and the insertion loss of the delay elements. Practical results of a feedforward system using class C main and error amplifiers are given.

Two methods to improve the power efficiency of the feedforward amplifier are presented. In the previous section, it was shown that it is possible to reduce the main path delay, and still retain significant linearity (particularly in systems with a small relative bandwidth). Reducing the main path delay (or even eliminating it), reduces its loss, and hence improves efficiency. An alternative method is to utilise the power from the error amplifier wasted in the output coupler to supplement the dc power supplied to the amplifiers. In highly nonlinear, high efficiency applications, this recovered power can improve the overall power efficiency by a significant amount.

4.3.2 Typical Theoretical Characteristics

The efficiency of a feedforward amplifier is theoretically analysed in Appendix B (Section B.2) and is given by

$$\eta_{ff} = \frac{\eta_{A1}\eta_{A2}C_{DC}(1 - C_{DC})}{\eta_{A1}F_{IM}(1 - C_{DC}) + \eta_{A2}LC_{DC}(1 + F_{IM})} \tag{4.4}$$

where η_{A1} and η_{A2} are the efficiencies of the main and error amplifiers respectively, F_{IM} is the power of the third–order IMPs relative to the linear output, C_{DC} is the coupling factor of C2 and L is the delay insertion loss (all in linear terms). The coupling factor which gives optimal efficiency is given by

$$C_{DC,OPT} = \frac{\eta_{A1}F_{IM} - \sqrt{\eta_{A1}\eta_{A2}F_{IM}L(1 + F_{IM})}}{\eta_{A1}F_{IM} - \eta_{A2}L(1 + F_{IM})} \tag{4.5}$$

The sensitivity of a particular function, y, to a parameter x is denoted as S_x^y; for example a sensitivity of 5 means that a 1% change in x leads to a 5% change in y. The sensitivity of the feedforward amplifier efficiency to the error amplifier efficiency, $S_{\eta_{A2}}^{\eta_{ff}}$, may be shown to be

$$S_{\eta_{A2}}^{\eta_{ff}} = \frac{\eta_{A1}F_{IM}(1 - C_{DC})}{\eta_{A1}F_{IM}(1 - C_{DC}) + \eta_{A2}LC_{DC}(1 + F_{IM})} \tag{4.6}$$

The analysis is illustrated using two theoretical feedforward amplifiers each having either class A or class C main and error amplifiers. The class A amplifiers are assumed to have power efficiencies of 15%, with third–order IMPs 35dB below the carrier level (−35dBc) in a two–tone test. The class C amplifiers are assumed to have power efficiencies of 60%, with third–order IMPs at −15dBc.

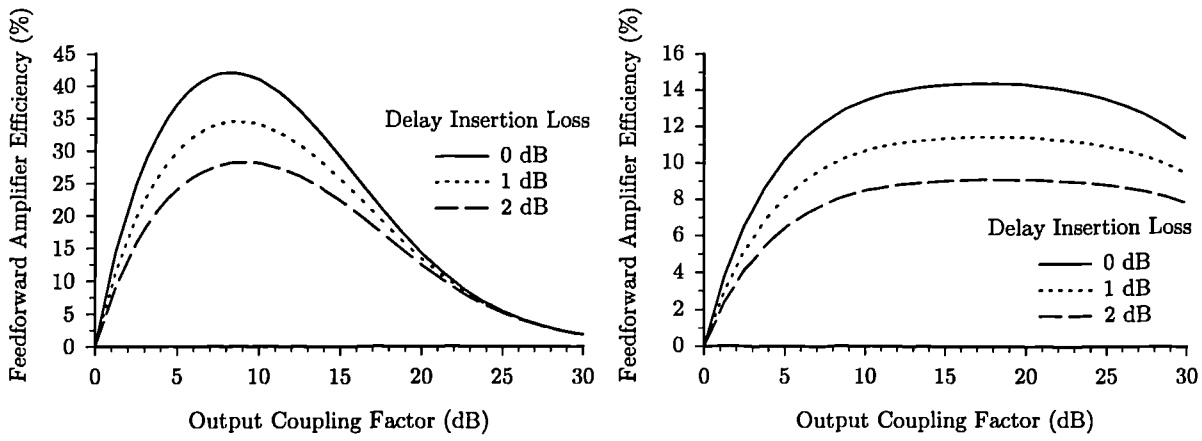
The efficiencies of the two feedforward amplifiers are compared in Figure 4.8. In the ideal case of zero delay insertion loss, the class C–based feedforward amplifier has a maximum efficiency of 42%. Whilst this is a significant reduction in efficiency compared to the main amplifier alone (60%), the amplifier still shows reasonable efficiency. The maximum efficiency of the class A–based feedforward amplifier assuming zero delay loss is 14.5%; a negligible reduction in main amplifier efficiency (15%).

The maximum efficiencies can be seen to be markedly reduced with even small amounts of delay insertion loss. In the class C case the maximum efficiency is reduced from 42% in the ideal case, to 35% with 1dB of delay insertion loss, and to only 28% with 2dB of loss. In the class A case, whilst there is negligible difference between the main amplifier and feedforward amplifier efficiencies with zero insertion loss, this is not the case if delay loss is present. The maximum efficiency is reduced from 14.5% in the ideal case to 11.5% and 9% with 1dB and 2dB of delay loss respectively.

Another effect of the delay insertion loss is that a higher power main amplifier is required to generate a given output power. For example, a delay insertion loss of 1dB would require the main amplifier to have a peak envelope power (PEP) rating approximately 25% greater than that required with zero delay loss.

The coupling factor value is more critical for the class C amplifier case than for the class A case. In the latter case, it is possible to reduce the coupling factor without significantly affecting the efficiency or the main amplifier power rating, but lowering the error amplifier power rating significantly.

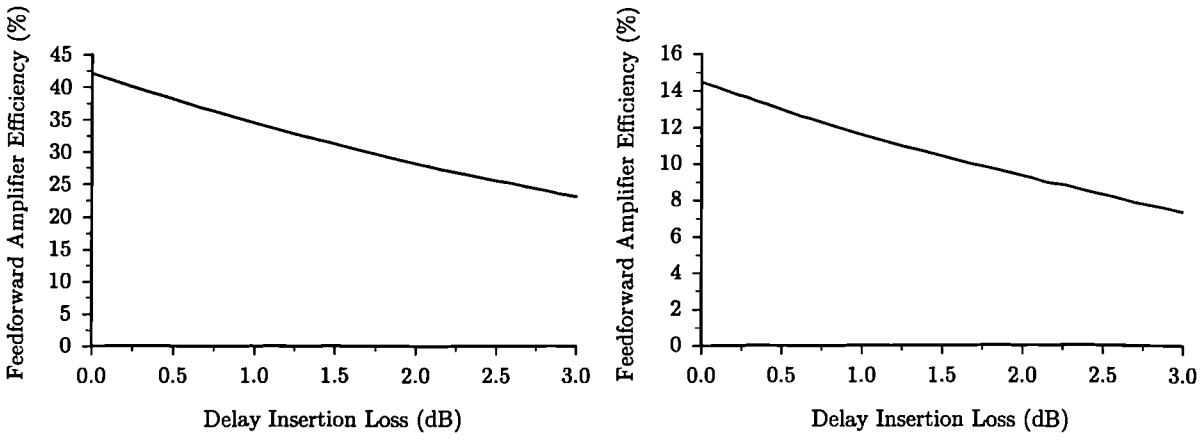
The maximum theoretical feedforward amplifier efficiency, assuming that the optimum coupling factor is used, is shown in Figure 4.9 for the two feedforward amplifiers. This illustrates how rapidly the feedforward efficiency falls as the delay insertion loss increases.



(a) Typical class C amplifiers

(b) Typical class A amplifiers

Figure 4.8: Feedforward amplifier efficiency using class C and class A amplifiers.



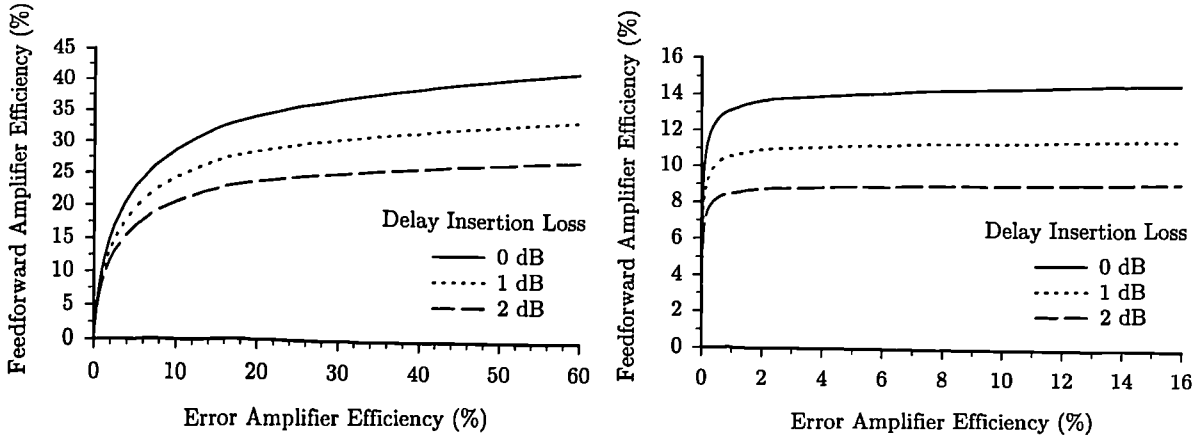
(a) Typical class C amplifiers

(b) Typical class A amplifiers

Figure 4.9: Maximum feedforward amplifier efficiency using class C and class A amplifiers (assuming optimal output coupling factor).

The feedforward amplifier efficiency is not highly dependent upon the error amplifier efficiency, as shown in Figure 4.10. This is to be expected since the error amplifier has a lower output power than the main amplifier, particularly if the main amplifier is reasonably linear, as is the case with the class A–based feedforward amplifier. Therefore, it may possible to increase the error amplifier linearity, at the expense of its efficiency, increasing the overall linearity without degrading overall efficiency to a large extent.

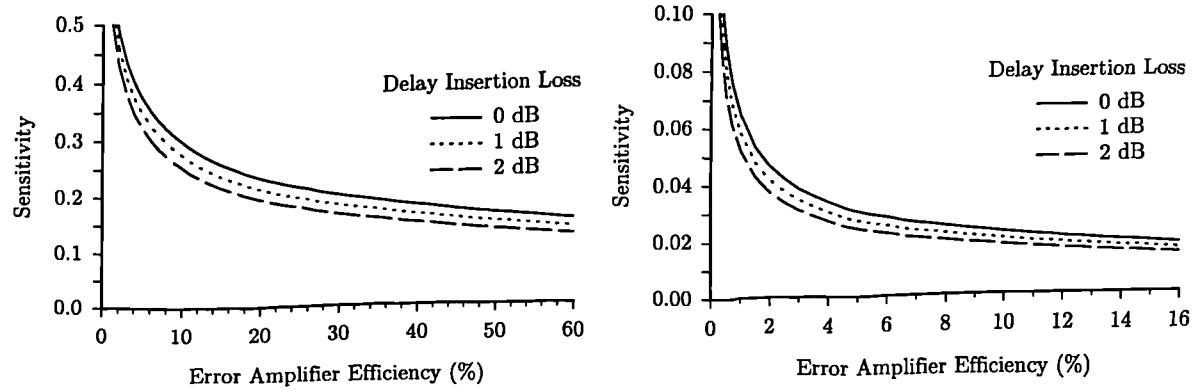
To illustrate this effect further, the sensitivity of the feedforward amplifier efficiency to the error amplifier efficiency with the class A– and class C–based feedforward amplifier examples are shown in Figure 4.11. In the case of the class C–based feedforward amplifier, reducing the error amplifier efficiency by 1% (from 60%) reduces the overall efficiency by approximately 0.14%, and in the case of the class A–based feedforward amplifier a 1% reduction (from 15%) causes approximately only a 0.015% reduction in overall efficiency.



(a) Typical class C amplifiers

(b) Typical class A amplifiers

Figure 4.10: Maximum feedforward amplifier efficiency versus error amplifier efficiency.



(a) Typical class C amplifiers

(b) Typical class A amplifiers

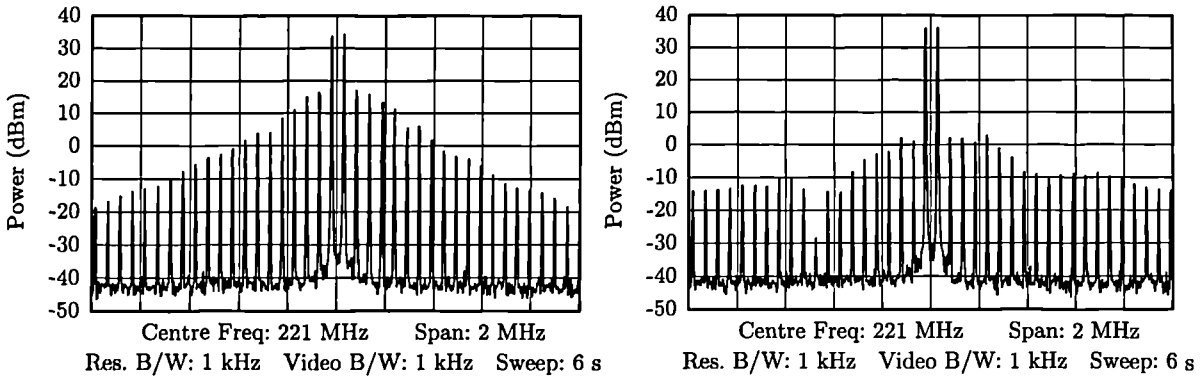
Figure 4.11: Sensitivity of feedforward amplifier efficiency (η_{ff}) to error amplifier efficiency (η_{A2}), $S^{\eta_{ff}}_{\eta_{A2}}$.

4.3.3 Practical Feedforward System Results

A feedforward system was constructed using high efficiency class C main and error amplifiers, with class A driver amplifiers. The main amplifier operated over the frequency range 220–222MHz, with a peak output of 10W (40dBm), and third–order IMPs nominally at −15dBc in a two–tone test. The coupling factors used were $C1 = 33.5$dB and $C2 = 8.9$dB; the time delays were implemented using coaxial cable, and the main path delay insertion loss was 1.0dB. The PEP rating of the error amplifier, with the output coupling factor of 8.9dB, was required to be greater than 2.5W (33.9dBm). The two class C amplifiers were both constructed using VMOS transistors which have a number of advantages over BJTs, particularly in that their gain and efficiency are higher, and that the bias point is much easier to adjust, allowing fine control of the efficiency and distortion.

Frequency Domain Results

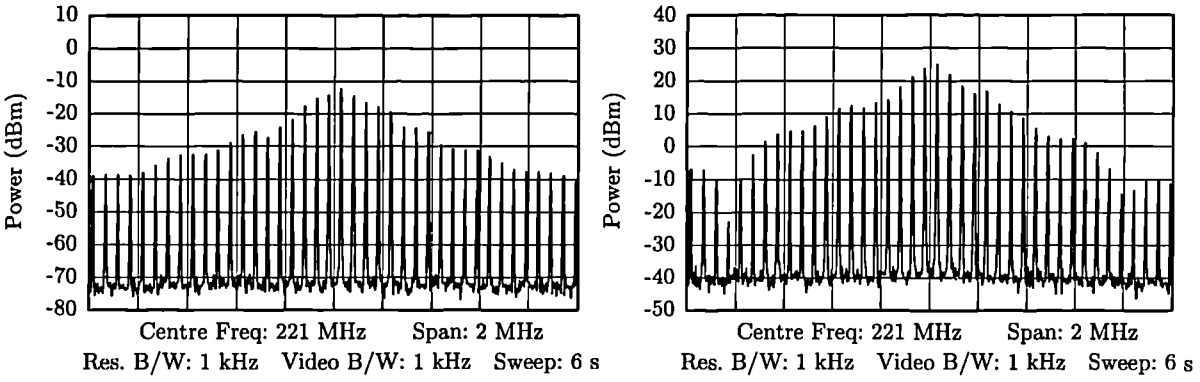
The uncorrected output signal of the main amplifier with a two–tone test input signal is shown in Figure 4.12(a). The two tones are spaced 50kHz apart in the centre of the band, and the output power per tone is 34dBm (corresponding to approximately 40dBm PEP). The third–order IMPs are at a level of −16dBc, and the other IMP orders are also at a significant level. The output of the feedforward amplifier is shown in Figure 4.12(b); the highest IMP level is approximately −35dBc, similar to that of a good class A amplifier.



(a) Uncorrected output

(b) Corrected (feedforward) output

Figure 4.12: Practical uncorrected and corrected amplifier output with two–tone input.



(a) Subtracter output

(b) Error amplifier output

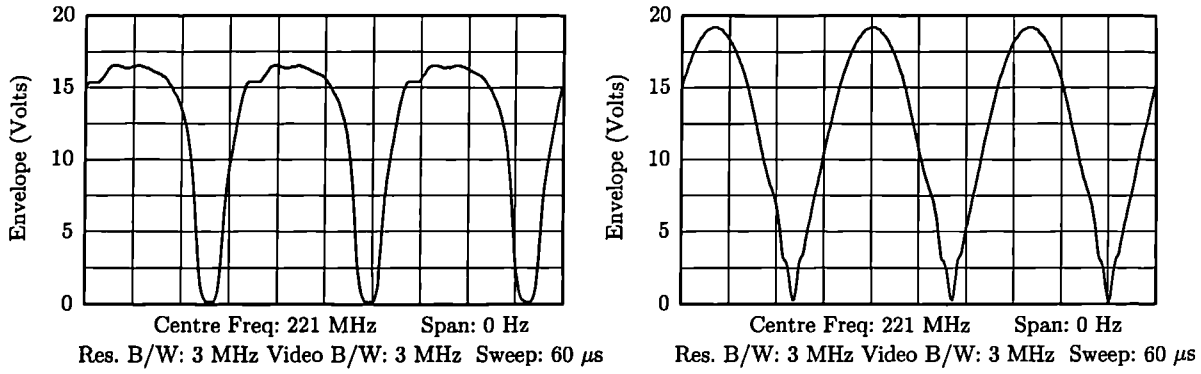
Figure 4.13: Output of subtracter and error amplifier with two–tone input.

To achieve the best distortion cancellation in the system output, it was found necessary to adjust the main loop amplitude and phase shift such that the original two tones were not optimally cancelled. The error signal at the output of the subtracter is shown in Figure 4.13(a), which is then amplified by the class C error amplifier to obtain the error signal shown in Figure 4.13(b). It can be seen that the two error signals are significantly different, due to the high degree of nonlinearity of the error amplifier. The residual two tones will intermodulate with the distortion whilst passing through the error amplifier, altering the characteristics of the IMPs at its output. The resultant IMPs have an amplitude and phase which is, therefore, dependent upon the nonlinearity of the error amplifier and upon the level of the original two tones. Thus it is possible to reduce the effect of the distortion introduced by the error amplifier on its output signal, by injecting more of the original carriers into the error loop.

In practice, an overall linearity improvement of more than 6dB has been achieved by this technique.

Time Domain Results

The envelope of the uncorrected output signal from the main amplifier is shown in Figure 4.14(a). There is a small region of the envelope in which the transistor is cut–off, followed by a very rapid rise, as the transistor starts to conduct, and finally saturation.



(a) Uncorrected output

(b) Corrected (feedforward) output

Figure 4.14: Envelope of uncorrected and corrected amplifier output with two–tone input.

The envelope of the feedforward amplifier output is shown in Figure 4.14(b). The cut–off and saturation regions have been eliminated, since the error amplifier supplies the output power under these conditions. The envelope has a much more gradual rise, consistent with the overall amplifier operating more linearly.

Power Efficiency Results

A two–tone input applied to the main amplifier produces an efficiency close to 55% for the main amplifier chain (*i.e.*, main amplifier plus driver amplifier), and an efficiency of approximately 40% for the error amplifier chain. The low error amplifier efficiency occurs because the peak–to–mean ratio of its input signal is much higher than that for two tones, due to the large number of distortion products present in its input spectrum. This effect has been examined in [9]. Using these efficiencies, and the measured delay insertion loss of 1.0dB, the predicted feedforward amplifier efficiency is 31%. The actual efficiency measured was 27%. The discrepancy is mainly due to the higher–order IMPs present, which have not been considered in the theoretical analysis.

4.3.4 Improving the efficiency by reducing the Main Path delay

The feedforward amplifier efficiency can be significantly improved by reducing the loss of the main path delay. This can be achieved by using high quality delay lines, or by reducing the time delay. The former solution is costly, and some loss is inevitable, creating a limited improvement. The latter solution will introduce delay mismatch between the two cancellation paths, having an effect on the linearity. This effect has been investigated in Section 4.2.1, and it was shown that significant levels of distortion cancellation are still possible, particularly if small relative bandwidths are to be used.

To practically demonstrate the effects of reducing the main path delay in a feedforward system, the feedforward amplifier described in the previous section was used. The main path delay was eliminated; equivalent to a delay mismatch of 5 cycles of the centre frequency. The output signal of this amplifier with a two tone test applied is shown in Figure 4.15. The maximum IMP level is −32dBc, compared with −35dBc for the response with the correct delay (Figure 4.12(b)). Thus, most of the linearity has been retained. However, it can be seen that the higher–order IMPs are not cancelled to the same extent, consistent with the fact that the cancellation is reduced across the band. The efficiency has been increased to 33% by removing the delay, from an original value of 27% (an increase of 6%).

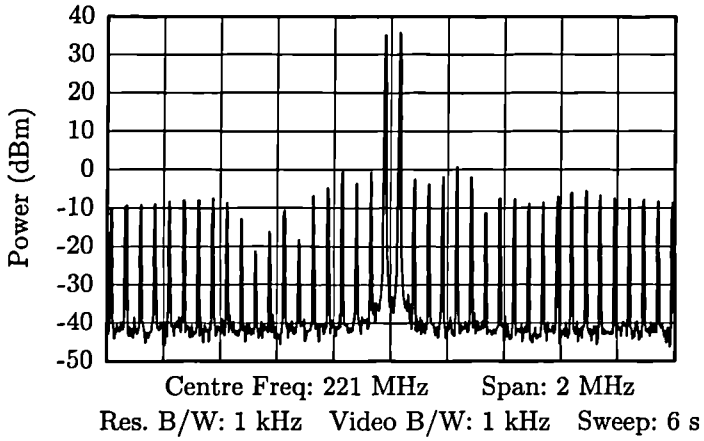


Figure 4.15: Feedforward amplifier output with two tone input and 5 cycles of main path delay mismatch.

4.3.5 Improving the efficiency by power recovery

One source of inefficiency in a feedforward amplifier is the error amplifier power wasted in the output coupler, $C2$. For example, assume a class C main amplifier generating 10W, with IMPs at −15dBc in a two–tone test. Using a 10dB output coupler would require an error amplifier power of approximately 3W. However, 90% of this power would be wasted in the output coupler termination, corresponding to 2.7W. One method of recovering the wasted power has been suggested in a LINC combiner, and consists of replacing the 50Ω termination of the coupler by a matched RF rectifier[1]. The rectified RF signal may be used to supplement the dc supply for example by charging the supply batteries. This could be done with high efficiency, recovering 70–85% of the wasted power. The new configuration is shown in Figure 4.16; the amplitude and phase controllers have been removed for clarity.

Typical Theoretical Characteristics

Theoretical analysis of the feedforward amplifier efficiency utilising this technique is presented in Appendix B (Section B.3). The efficiency of a feedforward amplifier with power recovery is given by

$$\eta_{ff} = \frac{\eta_{A1}\eta_{A2}C_{DC}(1-C_{DC})}{\eta_{A1}F_{IM}(1-C_{DC}) + \eta_{A2}LC_{DC}(1+F_{IM}) - \eta_{A1}\eta_{A2}\eta_{PR}F_{IM}(1-C_{DC})^2} \tag{4.7}$$

[1]The author acknowledges Dr. Peter Kenington as the originator of this technique.

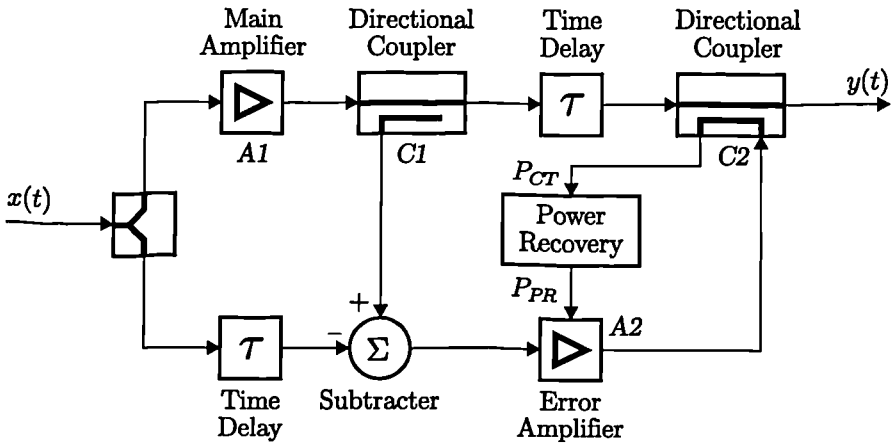
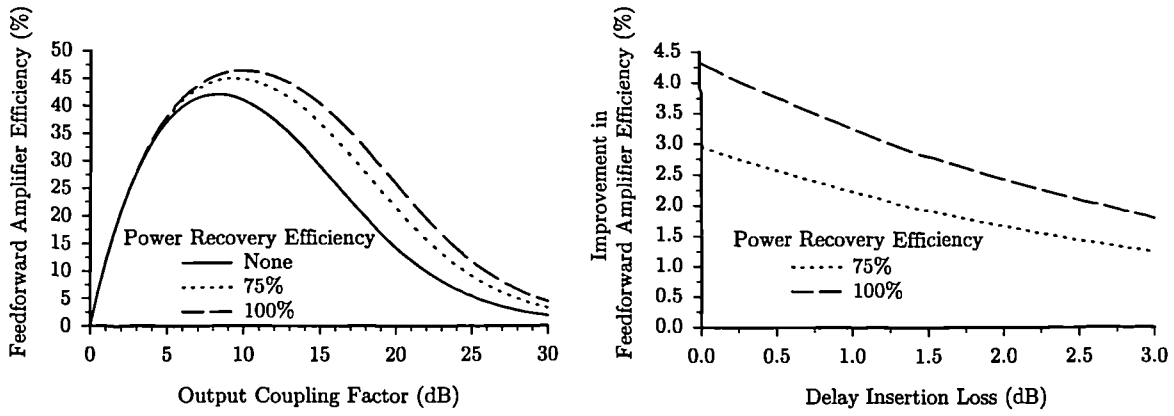


Figure 4.16: The configuration of a feedforward amplifier with power recovery.

where η_{PR} is the efficiency of the power recovery circuitry. The coupling factor which gives optimal efficiency is given by

$$C_{DC,OPT} = \frac{\eta_{A1}F_{IM}(1-\eta_{A2}\eta_{PR}) - \sqrt{\eta_{A1}\eta_{A2}F_{IM}L(1+F_{IM})(1-\eta_{A2}\eta_{PR})}}{\eta_{A1}F_{IM}(1-\eta_{A2}\eta_{PR}) - \eta_{A2}L(1+F_{IM})} \tag{4.8}$$

The improvement in efficiency which can be provided by power recovery for the class C–based feedforward amplifier is shown in Figure 4.17. The efficiency could be improved by 4.3% with a 100% power recovery efficiency, or by 3% with a 75% power recovery efficiency, assuming no delay insertion loss. The improvement which can be attained is reduced as the delay insertion loss increases, and depends greatly upon the error amplifier power, and hence the linearity of the main amplifier. Therefore, using the technique with class A amplifiers does not give any significant efficiency improvement.



(a) Feedforward amplifier efficiency

(b) Improvement over standard feedforward

Figure 4.17: Feedforward amplifier efficiencies using standard and power–recovered system topologies with typical class C amplifiers.

4.4 Combined use of Predistortion and Feedforward Linearisation

The efficiency of a feedforward amplifier has been shown in the previous section to be significantly lower than that of the main amplifier if the main amplifier is nonlinear. The efficiency can be improved by some extent using the techniques described. An alternative method is now described which can improve both the linearity and the efficiency of a feedforward amplifier by a significant amount.

Analogue predistortion techniques, described in Section 3.6.3, can currently improve the linearity of an amplifier by only a small degree (typically 10–15dB), whilst retaining the amplifier power efficiency[2]. This modest level of linearity improvement prohibits its exclusive use in most applications requiring good linearity. However, the use of predistortion on the main amplifier in a feedforward system could result in a significant improvement in overall efficiency and linearity over feedforward linearisation alone. The error signal would be at a much lower level, since the main amplifier has been made more linear due to the use of predistortion, and hence the error amplifier power rating can be much reduced. This allows the use of a more linear (and less efficient) error amplifier, without degrading the overall efficiency, whilst improving overall linearity. The effect of predistortion of the main amplifier on the overall efficiency of the theoretical class C–based feedforward amplifier is shown in Figure 4.18 (this figure has been obtained using Equation 4.4 by reducing the initial IMP level from −15dBc by the predistortion level). Without predistortion, the maximum efficiency which can be attained is 42%; this is significantly increased by using only small amounts of predistortion. With 10dB of predistortion improvement, the efficiency is 52%, and with 20dB of predistortion improvement it is 57%. The main amplifier efficiency is 60% and, therefore, the reduction in efficiency caused by the use of feedforward is very small.

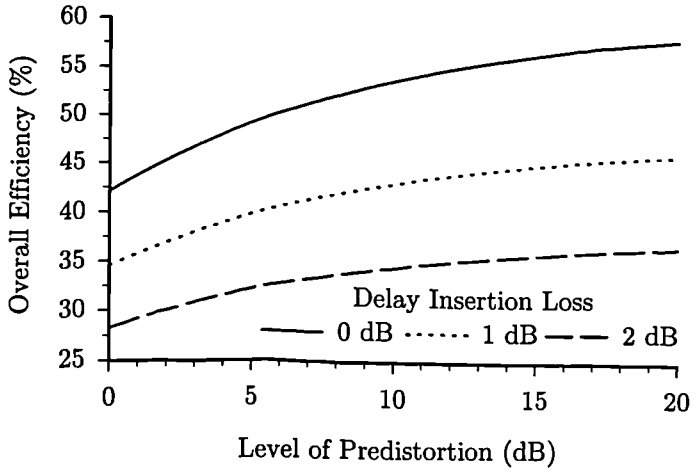


Figure 4.18: Overall Efficiency of a Feedforward Amplifier with Predistortion using class C amplifiers.

The main path delay loss still affects the overall efficiency. However, the improved linearity of the main amplifier due to predistortion could enable the use of more delay mismatch, whilst still retaining linearity.

The sensitivity of the feedforward amplifier efficiency to the error amplifier efficiency with the

[2]However, the author knows of no published material for analogue predistortion of class C amplifiers. In such cases the linearity improvement with current forms of analogue predistortion would be expected to be reduced.

predistorted theoretical class C–based feedforward amplifier example, is shown in Figure 4.19. It can be seen that increased levels of predistortion rapidly reduce the sensitivity of the feedforward amplifier efficiency to the error amplifier efficiency. Thus, it becomes acceptable to use a more linear error amplifier to improve the linearity, without its reduced efficiency significantly degrading the overall efficiency.

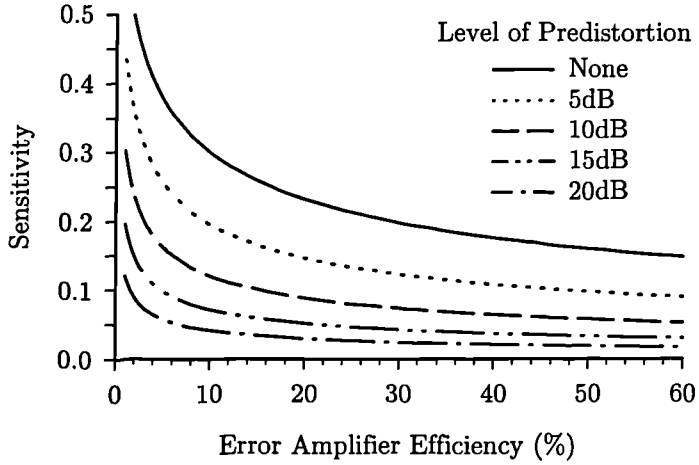


Figure 4.19: Sensitivity of the class C–based feedforward amplifier efficiency (η_{ff}) to error amplifier efficiency (η_{A2}), $S_{\eta_{A2}}^{\eta_{ff}}$, with predistortion of the main amplifier, assuming no main path delay loss.

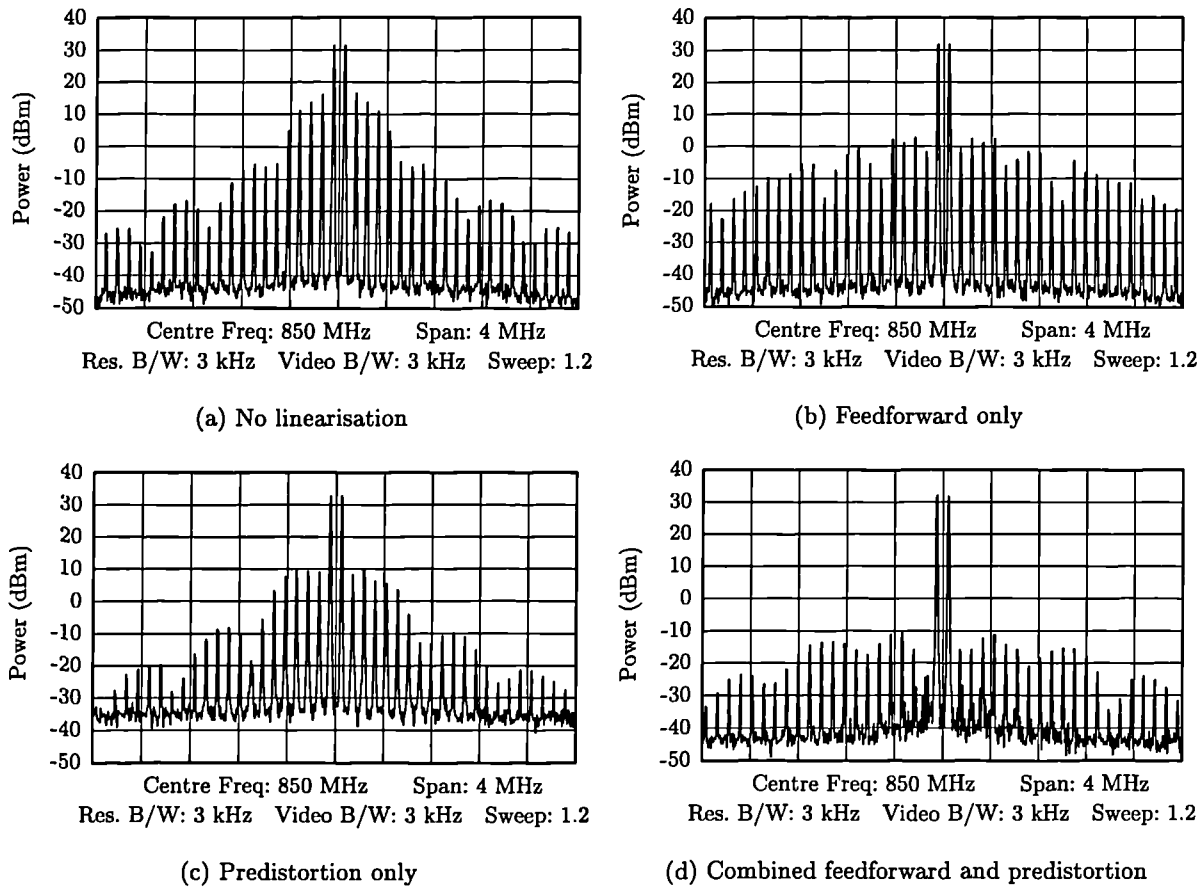
4.4.1 Practical Results

Another practical amplifier system has been constructed, operating at a centre frequency of 850MHz. This has been linearised using both feedforward and predistortion individually, and by a combination of both techniques. The main amplifier used is a class C power module, the output spectrum of which is shown in Figure 4.20(a) when a two–tone signal is applied (at 100kHz spacing). The highest IMP level is at −14dBc at an output power of 33dBm per tone. The power efficiency under these conditions is 35%.

Feedforward linearisation only

Feedforward linearisation of this amplifier has been applied, using another class C power module as the error amplifier, with the delay elements removed. The main path loss of 1.5dB is made up of the through loss of the coupler (0.65dB) along with further losses introduced by connectors and mismatches, and reduces the level of the output tones to 31.5dBm. The output spectrum of the feedforward amplifier with a two–tone signal applied is shown in Figure 4.20(b). The maximum IMP level has been reduced to −30dBc, an improvement of 16dB.

The error amplifier power rating is required to be approximately 20% of that of the main amplifier. This is a very significant fraction of the main amplifier rating, and thus the error amplifier will have a large effect on the overall efficiency. The two–tone power efficiency of the feedforward amplifier is 19%, compared with 35% for the main amplifier alone. The maximum theoretical efficiency with the output coupling factor of 8.5dB, and the measured error amplifier chain efficiency of 21.5%, may be calculated using Equation 4.4, yielding 21%.



(a) No linearisation

(b) Feedforward only

(c) Predistortion only

(d) Combined feedforward and predistortion

Figure 4.20: Amplifier output signals with two–tone input applied, with and without linearisation.

Analogue predistortion only

An intermediate frequency (IF) scalar analogue predistorter was constructed, operating at a centre frequency of 200MHz. The block diagram for the predistorter is shown in Figure 4.21 (the figures are unordered to facilitate comparison of the output spectra). The output spectrum of the predistorted amplifier with a two–tone test signal applied, operating at an output power of 33dBm per tone, is shown in Figure 4.20(c). The performance of a predistortion amplifier will be more thoroughly examined in the following chapter.

The maximum IMP level of the amplifier has been reduced by 9dB to −23dBc. Whilst this is substantially less than the improvement using feedforward linearisation, the power efficiency has been largely unaltered (32.5%), and the output power is slightly greater due to the lack of main path couplers. The extra power consumption introduced by the IF strip has been ignored, to allow meaningful comparison of the techniques, and because it would have less effect on the overall efficiency if a higher power main amplifier were to be used. The linearity improvement is relatively modest, and illustrates the difficulties in predistorting class C amplifiers.

Referring to the predistortion amplifier output spectrum, it can be seen that there is a significant increase in the noise floor. This is due to the predistorter network and the mixers adding noise to the amplifier input.

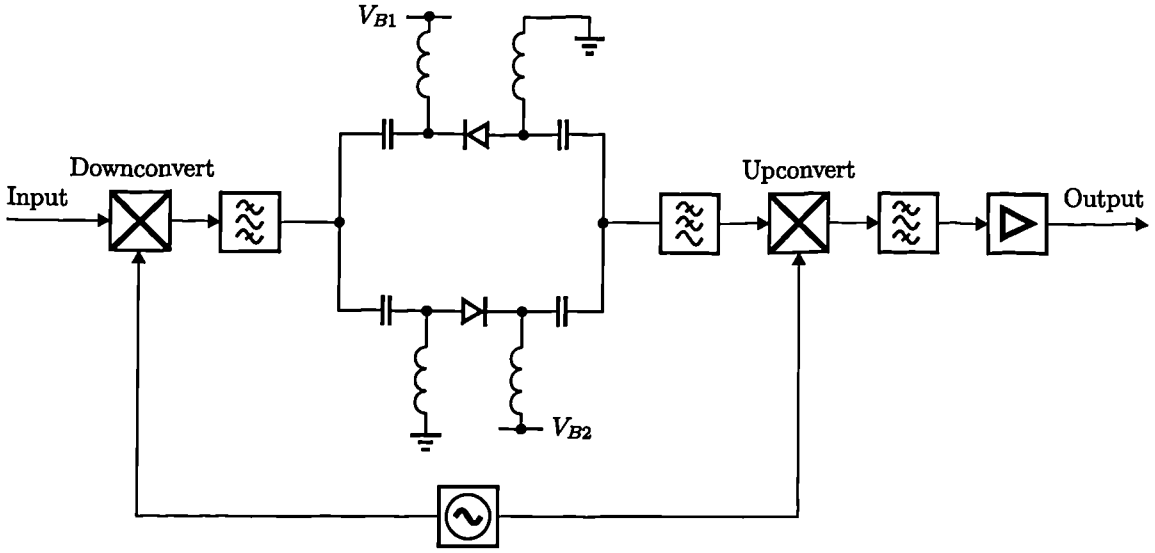


Figure 4.21: Block diagram of the constructed IF predistorter.

Combined predistortion and feedforward linearisation

To investigate the performance of the combined use of feedforward and predistortion, the main amplifier in the feedforward system is predistorted, and a class A error amplifier used. The predistorted main amplifier has an IMP level of −23dBc, and with the chosen error injection coupler value of 10dB, the PEP rating of the error amplifier has been substantially reduced to approximately 5% of the main amplifier rating.

There is a small loss through the error injection coupler (ideally 0.4dB) and, together with other losses, the total main path insertion loss is 0.8dB. This additional loss reduces the output power and the overall efficiency to some degree. Without the feedforward correction, but with the couplers connected, the efficiency of the predistorted amplifier is reduced from 32.5% to 28%.

The two–tone test output spectrum of the combined predistortion and feedforward amplifier is shown in Figure 4.20(d); the highest IMP level is at −43dBc, an improvement of 29dB. The overall two–tone power efficiency is 21%. The predicted overall efficiency assuming no extra main path losses (other than the error injection coupler) is 28%; this is reduced to 25% with the measured additional loss.

The efficiency is slightly greater than that attained using feedforward linearisation alone (21% as opposed to 19%), although the theoretical efficiency is significantly greater (28% as opposed to 21%). Thus, there is much room for improvement of the combined amplifier system in terms of its efficiency. The linearity improvement is much greater than that attained using either linearisation technique alone (29dB compared with 16dB for feedforward linearisation, and 9dB for predistortion). Both the efficiency and the linearity can be increased by improving the predistortion network, which currently only provides a reduction of distortion of 9dB, to produce a highly linear and efficient amplifier.

In the combined predistortion and feedforward case, feedforward linearisation is providing 20dB of distortion cancellation. It is anticipated that the linearity may be further improved, without significant loss in overall efficiency, by improving the linearity of the error amplifier. It is likely that the distortion cancellation provided by the feedforward technique could be increased to better than 30dB, resulting in IMPs of less than −55dBc. The linearity may be increased still further by improvements in the predistortion network.

Referring to the output spectrum of the combined predistortion and feedforward amplifier, it can be seen that the noise introduced by the predistorter, evident in Figure 4.20(c), has been significantly reduced. This is due to the fact that the noise is added into the main path, and does not appear in the error path; thus, it will be cancelled in the same manner as the distortion. This is an added advantage of applying feedforward linearisation to a predistortion amplifier.

4.5 Summary

The linearity of a feedforward amplifier has been theoretically examined, and it has been shown that to obtain good linearity requires very close amplitude and phase balance of the two paths across the band. To maintain this balance will, in general, require the use of an adaptive control mechanism. A scheme has been described which reduces the frequency–dependent amplitude and phase variation across the band, potentially providing increased distortion cancellation.

In applications with large relative bandwidth (> 3%), it is necessary for the gross phase shift to be as close to 180° as possible, with no delay mismatch, in order to achieve high levels of cancellation. In small relative bandwidth applications this is not so important, since the amplitude and phase errors caused by component frequency–dependency, and non–optimal adaption, will usually have a more significant effect. In these applications it may also be acceptable to reduce the delay times by integer numbers of cycles of the centre frequency. Whilst this will reduce linearity to some extent, the increase in power efficiency is significant. The degradation in linearity so caused may be insignificant due to the other amplitude and phase errors, as previously described.

The efficiency of a feedforward amplifier has been examined, and the effects of main path delay insertion loss and output coupling factor have been described. It has been shown that there is an optimal coupling factor to attain highest efficiency, and that its value is more critical when highly nonlinear main amplifiers are used. The theory presented also shows that minimisation of the main path delay insertion loss in a feedforward linearisation system is critical in high–efficiency applications. Failure to do this results in severely reduced efficiency, and a significant increase in the required power rating of the main amplifier. The theoretical results presented give good agreement with those of the practical feedforward amplifier.

In the practical class C amplifier, the linearity of the main amplifier has been significantly improved by feedforward linearisation, with a reduction in the highest IMP level of 19dB. The third–order IMPs were reduced to a level approximating that of a class A amplifier. The use of a nonlinear error amplifier causes the output to have significant levels of high–order distortion, which does not occur in class A amplification, and increases its spectral occupancy. The effect which the error amplifier nonlinearity has on the overall linearity performance can be counteracted to some extent by retaining a significant level of the original two tones. The efficiency of the feedforward amplifier is significantly better than that attainable with class A amplifiers of similar linearity. It is anticipated that this could be further increased with the use of improved driver amplifiers.

The high–order distortion produced limits the applicability of using class C–based feedforward amplifiers to systems requiring only modest linearity, such as satellite transmitters, or those involving a small envelope variation on the RF signal (such as certain filtered digital modulation formats). To counter this, additional feedforward processes may be added to improve linearity at the expense of power efficiency and complexity, or a combination of linearisation schemes used.

The efficiency of the practical feedforward amplifier was improved by eliminating the main path delay. This was found to improve the efficiency by 6%, whilst only increasing the level of the highest IMP by 3dB.

Theoretical analysis was also presented for a scheme which could be used to improve the efficiency of feedforward amplifiers by making use of the power from the error amplifier which is lost in the output coupler. The effect has been shown to only be significant when highly nonlinear main amplifiers are used, with very small delay insertion loss. In practice, therefore, it is unlikely that the extra complexity added by the use of the power recovery technique, to gain a small improvement in efficiency, would warrant its use.

Finally, it has been suggested that, to counter the poor power efficiency characteristics of a feedforward amplifier, whilst also improving linearity, a combination of predistortion and feedforward linearisation be used. To illustrate this, a practical system has been constructed using either predistortion or feedforward linearisation alone, and then a combination of the two. It has been shown that the combined system can offer both improved linearity and efficiency compared with either technique in isolation. However, at present, the linearity improvement of analogue predistortion techniques is small, which limits the performance of the combined system. Therefore, further study of analogue predistortion is required, and will be presented in the following chapters.

REFERENCES

[1] R GERARD. *UK Patent Application GB 2 080 062 A*, January 1982.

[2] R. MYER. *U.S. Patent 4,580,105*, April 1986.

[3] R. BAUMAN. *U.S. Patent 4,389,618*, June 1983.

[4] J. POWELL. *UK Patent GB 2 238 196 A*, May 1991.

[5] P. B. KENINGTON, M. A. BEACH, A. BATEMAN, AND J. P. MCGEEHAN. *U.S. Patent 5,157,345*, October 1992.

[6] D.W. BENNETT. *The design and performance of automatically-controlled feedforward amplifiers.* PhD thesis, Dept. of Electrical & Electronic Engineering, University of Bristol, September 1995.

[7] P.B. KENINGTON AND D.W. BENNETT. Linear distortion correction using a feedforward system. *IEEE Transactions on Vehicular Technology*, VT-45(1):74–81, February 1996.

[8] P. B. KENINGTON. Efficiency of feedforward amplifiers. *IEE Proceedings-G*, 139(5):591–593, October 1992.

[9] F.H. RAAB. Average efficiency of power amplifiers. In *Proc. of RF Technology Expo 86*, pages 473–486, Anaheim, CA, USA, February 1986. Cardiff Publishing Co.

Chapter Five

Broadband Predistortion Linearisation

This chapter describes the factors which affect the linearity and efficiency of a predistortion amplifier. The ideal predistorter characteristic is derived, and a computer simulation is presented to predict the performance of a predistortion amplifier. The instantaneous and average efficiencies of a predistortion amplifier are derived, and using this information its average efficiency using various representative input signals is calculated, and compared with that of the amplifier alone.

5.1 Introduction

Broadband linearisation of RF power amplifiers can be achieved with either the feedforward technique described in Chapter 4, or the predistortion technique. Broadband predistortion generally uses analogue components, operating at either the carrier frequency (RF predistortion), or some intermediate frequency (IF predistortion). The linearity and efficiency characteristics of a predistortion amplifier will be further examined in this chapter.

5.2 Linearity Considerations of Predistortion Linearisation

5.2.1 Ideal Predistorter Characteristic

The configuration of a predistortion amplifier is shown in Figure 5.1. The transfer characteristic of the predistorter is chosen to minimise the level of distortion at the output of the amplifier. It is possible to deduce the required predistorter transfer characteristic from that of the amplifier in the following manner.

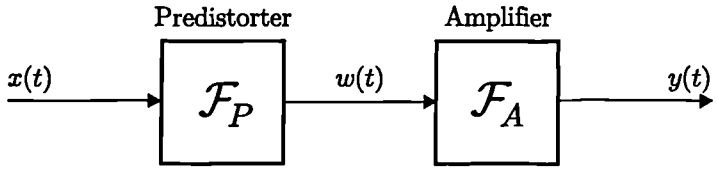


Figure 5.1: Configuration of a predistortion amplifier.

An input signal, $x(t)$, is applied to the predistorter which has a transfer characteristic, $\mathcal{F}_P(\cdot)$, generating an output signal, $w(t)$, thus

$$w(t) = \mathcal{F}_P(x(t)) \tag{5.1}$$

This is then applied to the amplifier, and generates the output $y(t)$, where

$$y(t) = \mathcal{F}_A(w(t)) \tag{5.2}$$

It may be easily seen that the amplifier output in terms of the overall (predistorter) input is given by

$$y(t) = \mathcal{F}_A\big(\mathcal{F}_P(x(t))\big) \tag{5.3}$$

Therefore, if the output $y(t)$ is required to be a linearly amplified version of the overall input $x(t)$, *i.e.*, $y(t) = kx(t)$, the predistorter characteristic is required to be

$$\mathcal{F}_P(x(t)) = \mathcal{F}_A{}^{-1}(kx(t)) \tag{5.4}$$

Thus, the ideal predistorter characteristic is given by the inverse of the amplifier transfer characteristic.

The ideal predistorter characteristic in terms of its AM/AM and AM/PM conversion is now determined. This allows simple determination of the ideal predistorter characteristic from easily measured amplifier data. The input signal, $x(t)$, in analytical form, is assumed to be a modulated carrier

$$x(t) = A(t)e^{j(\omega t + \phi(t))} \tag{5.5}$$

To avoid confusion, time will be made implicit for the modulation. Hence,

$$x(t) = Ae^{\jmath(\omega t+\phi)} \tag{5.6}$$

This is applied to the predistorter which has AM/AM conversion $F_P(\cdot)$ and AM/PM conversion $G_P(\cdot)$ to give an output $w(t)$, thus

$$w(t) = F_P(A)e^{\jmath G_P(A)}e^{\jmath(\omega t+\phi)} \tag{5.7}$$

The amplitude of the predistorter output signal, $|w(t)|$, is given by

$$|w(t)| = F_P(A) \tag{5.8}$$

This signal is then applied to the amplifier which has AM/AM conversion $F_A(\cdot)$, and AM/PM conversion $G_A(\cdot)$, to generate the output signal $y(t)$, thus

$$y(t) = F_A\big(F_P(A)\big)e^{\jmath(G_A(F_P(A))+G_P(A))}e^{\jmath(\omega t+\phi)} \tag{5.9}$$

To achieve linear gain k with a constant phase shift θ, the required output is given by

$$y(t) = kx(t)e^{\jmath\theta} \tag{5.10}$$

Hence,

$$kA = F_A\big(F_P(A)\big) \tag{5.11}$$

$$e^{\jmath\theta} = e^{\jmath(G_A(F_P(A))+G_P(A))} \tag{5.12}$$

Therefore, the predistorter AM/AM and AM/PM conversions are required to be given by

$$F_P(A) = {F_A}^{-1}(kA) \tag{5.13}$$

and

$$G_P(A) = \theta - G_A\big(F_P(A)\big) \tag{5.14}$$

or

$$G_P(A) = \theta - G_A\big({F_A}^{-1}(kA)\big) \tag{5.15}$$

5.2.2 Nonlinear Elements used for Predistortion

There are two general types of nonlinear elements used for the predistortion block; the polynomial predistorter and the generic predistorter.

Polynomial predistorter

The polynomial predistorter uses separate nonlinearities to reduce each order of distortion. Hence, a cubic nonlinearity can be used to reduce the third–order IMPs, and a quintic nonlinearity can be used to reduce fifth–order IMPs, *etc.*. These nonlinear outputs are added to a linear version of the input signal, and applied to the amplifier. However, the individual nonlinearities will not only affect the level of IMPs of the same order, but also lower orders (including the linear part).

The complexity of this type of predistorter usually limits its use to third–order IMP reduction. This will generally only be acceptable for amplifiers with predominantly third–order distortion, such as travelling wave tube amplifiers (TWTAs), and saturating class A solid–state amplifiers.

Generic predistorter

The generic predistorter uses a nonlinear device exhibiting a transfer characteristic which is the approximate inverse of the amplifier transfer characteristic. Since analogue predistortion appears to have thus far only been considered for saturating, reasonably linear amplifiers, the predistorter characteristics are generally expansive to counteract the compressive nature of the amplifier. Typical devices used include dual gate GaAs FETs operating close to pinch–off [1], or Schottky diodes [2].

5.2.3 Predistortion Architecture

Scalar predistorter

This predistortion architecture uses a single nonlinear element as the predistorter, of either the polynomial or generic type. Figure 5.2 shows the use of a scalar cubic polynomial predistorter, and its effect on the third–order distortion produced by the amplifier. The gain and phase of the paths are adjusted to achieve maximum cancellation of the distortion. It is not possible for a single nonlinearity to completely cancel the distortion introduced by the AM/AM and AM/PM conversion of the amplifier (unless only a single order of distortion is present), and so, performance is limited with this technique.

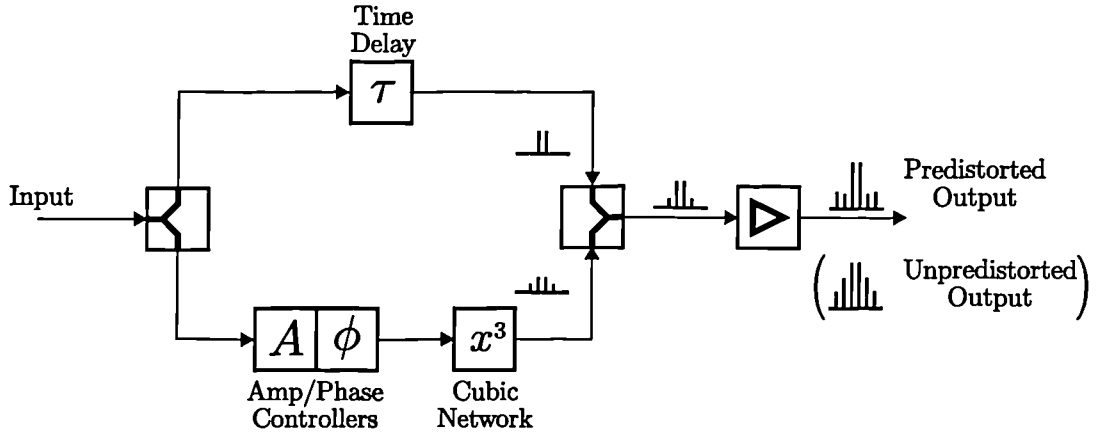


Figure 5.2: Scalar predistorter using a cubic nonlinearity.

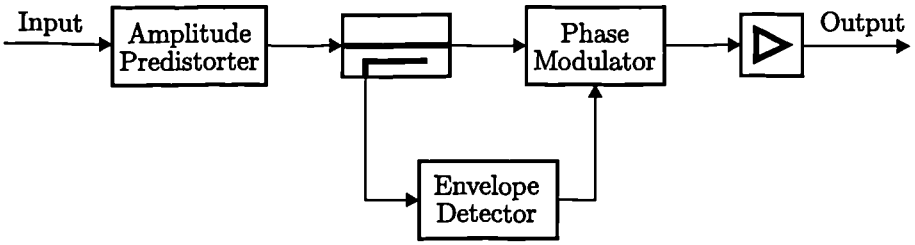
There have been many results published on the use of a scalar predistorter at IF and RF. Aihara *et al* [3] attained an improvement in third–order IMPs of 10dB at 6GHz using a cubic polynomial RF predistorter. Blair *et al* [4] obtained 8dB of suppression at the same centre frequency over a 30MHz bandwidth with a cubic polynomial IF predistorter.

Vector predistorter

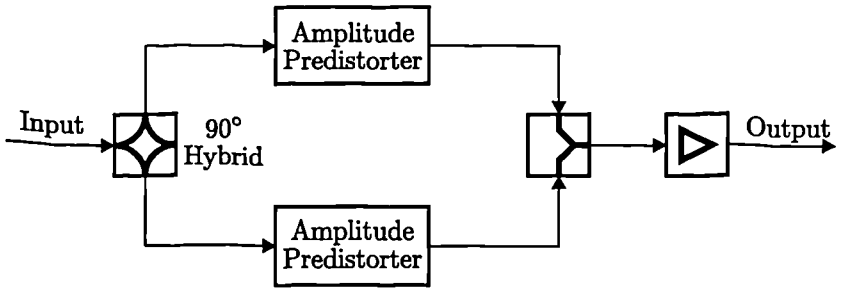
The distortion produced by the AM/AM and AM/PM conversion is orthogonal, and hence, two nonlinear elements, operating on orthogonal versions of the input signal, are required to eliminate it. The vector predistortion architecture applies either the polar (amplitude and phase) or quadrature components of the input signal to the nonlinear elements. The block diagram of both forms of the vector predistorter are shown in Figure 5.3. The vector predistorter has the same form as the memoryless bandpass nonlinear model described in Section 2.4.

The polar predistorter reduces the nonlinearity level by predistorting the envelope, and controlling the phase of the local oscillator providing the upconversion to the carrier frequency, via a phase modulator. The technique has attracted little attention, probably due to its increased complexity compared with the quadrature predistorter. Egger *et al* [2] achieved a reduction in third-order IMPs by 25dB, using 7 tones centred at 11.84GHz, with a bandwidth of 120MHz.

The quadrature predistorter first splits the input signal, via a 90° hybrid, into two quadrature paths, and uses an amplitude nonlinearity in each path to provide reduction of the AM/AM and AM/PM conversion. This technique has a high degree of symmetry, and is the preferred method of producing a vector predistorter. A quadrature predistorter has been used by Kumar *et al* [1] to achieve a 7dB reduction in third–order IMPs at a centre frequency of 12GHz, across a 500MHz bandwidth.



(a) Amplitude and phase predistorter



(b) Quadrature predistorter

Figure 5.3: Vector predistorter configurations.

5.2.4 Adaptive Control of a Predistorter

The predistorter architectures presented can be used to achieve small improvements in the linearity of an amplifier. To achieve optimal performance, it is necessary to use some form of adaptive control of the predistorter, due to the open–loop nature of the predistortion technique. However, since only modest linearity improvements are obtained, the predistorter is not generally very sensitive, and thus the adaption process is not critical. The adaption mechanisms thus far published [5–7] only adapt very slowly, and so can be used to counter the effects of temperature changes or ageing of components. More rapid changes in the amplifier characteristics may arise from sudden changes in loading, or operating frequency; these would necessitate a fast adaption mechanism. To the author's knowledge no rapid adaption techniques for analogue predistortion have thus far been proposed in the literature.

5.3 Simulation of Predistortion Amplifier Linearity

The ideal predistorter characteristic may not generally be found analytically, but can be simply found from measured amplifier data. This section describes a computer simulation which is used to characterise a predistortion amplifier using an ideal predistorter, and is to be used further in this work to analyse a predistortion amplifier with non–ideal predistorter networks.

The simulation blocks all use complex envelope equivalent forms, since this allows simple calculation of the signal envelope, and only produces a first harmonic zone output when amplified, as described in Section 2.4.3 (page 12). This greatly reduces the sampling rate required throughout the system, since no harmonics are produced, which would otherwise cause aliasing problems.

5.3.1 Test Signals

Two types of test signal, in complex envelope form, are used to test the performance of the predistortion amplifier. These are the two–tone signal, used to generate the spectral response of the system blocks; and the ramp signal, which is used to measure their transfer characteristics.

Two–Tone signal

The two–tone signal, $x[i]$, at iteration i with frequencies of f_1 and f_2, and sampling frequency of F_S is given by

$$\Re\{x[i]\} = \frac{1}{2}\left\{\cos\left(\frac{2\pi i f_1}{F_S}\right) + \cos\left(\frac{2\pi i f_2}{F_S}\right)\right\} \tag{5.16}$$

$$\Im\{x[i]\} = \frac{1}{2}\left\{\sin\left(\frac{2\pi i f_1}{F_S}\right) + \sin\left(\frac{2\pi i f_2}{F_S}\right)\right\} \tag{5.17}$$

where $\Re\{x[i]\}$ and $\Im\{x[i]\}$ are the real and imaginary parts of $x[i]$ respectively. The range of the signal is ± 1V, and has a peak amplitude of 1V.

Ramp signal

The ramp signal, $x[i]$, at iteration i with a frequency of f and sampling frequency of F_S is given by

$$\Re\{x[i]\} = \frac{2f}{F_S}\left\{i \,\%\, \left(\frac{F_S}{f}\right)\right\} - 1 \tag{5.18}$$

$$\Im\{x[i]\} = 0 \tag{5.19}$$

where % is the modulus operator (*i.e.*, it returns the remainder of the division). This signal has a range of ± 1V and the amplitude linearly increases from 0 to 1, with a constant phase of 0°. It can therefore be used to measure the AM/AM and AM/PM conversion of the system.

5.3.2 Amplifier Model

The amplifier model chosen for simulation is a bandpass, memoryless model, and thus has no frequency–dependence. The block diagram for the amplifier model is shown in Figure 5.4, and

is equivalent to the complex envelope form of the amplitude and phase model of a memoryless nonlinearity (Section 2.4.3). The RF signal is divided by the envelope to produce a signal consisting purely of the phase information. The envelope is then applied to the nonlinear function blocks, containing the AM/AM and AM/PM conversion information, via a voltage to dBm transformation. The AM/AM and AM/PM outputs of the nonlinear function blocks are converted into the required envelope and phase shift, and then multiplied by the phase information of the input signal to produce the nonlinearly amplified output.

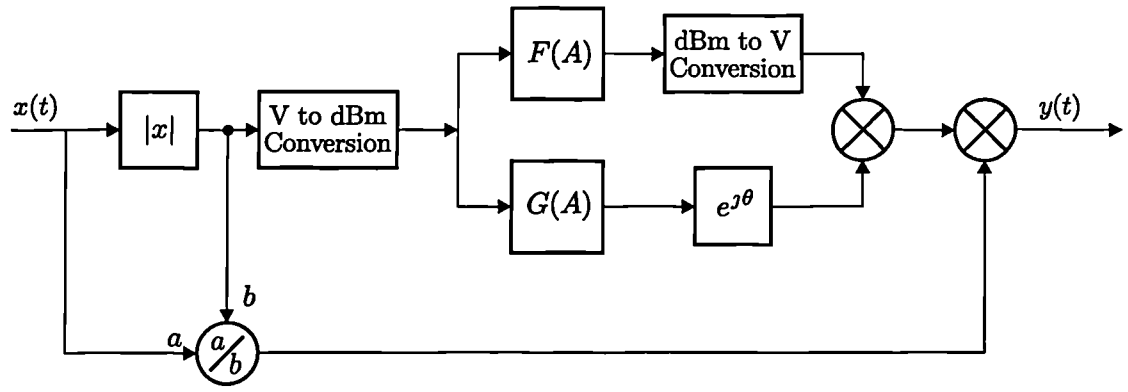


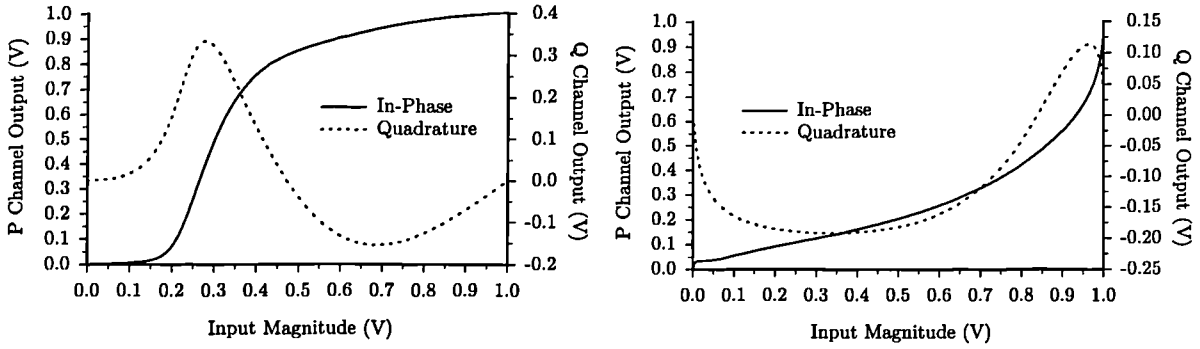
Figure 5.4: Block diagram of the amplifier model.

Due to the highly nonlinear nature of the amplifier to be modelled, the use of an approximating function of the forms described in Section 2.7 must be carefully considered. Polynomial approximation would require an impractical number of terms to accurately model the nonlinearity (it has been found that greater than 30 terms are required). Saleh functions, whilst able to accurately model weak nonlinearities due to amplifier compression, are unable to approximate class C amplifier nonlinearities with accuracy. Rational function and continued fraction approximation would require less terms than polynomial approximation, but a significant number would still be required, and fitting the function to the nonlinear data is a highly complex procedure. The most applicable approximating function is therefore a cubic spline, fitted to the data using interpolation. Interpolation ensures that the approximating function passes through all the data points, and so has no error at these points. However, this has the disadvantage that statistical errors (*e.g.*, noise) are not reduced, as would be the case with least–squared approximation using one of the other functions.

The nonlinear function blocks used to simulate the AM/AM and AM/PM conversion of the amplifier are formed using a look–up table of data measured from the class C main amplifier employed in the feedforward system described in Section 4.3.3. Cubic spline interpolation is used to approximate non–measured values, whilst data outside the measured range (below −40dBc) is determined by linear extrapolation. To improve the efficiency of the look–up process, the spline interpolation data is only calculated once, before the first iteration, and an efficient algorithm used to hunt through the look–up table. The measured data has been normalised such that a 0dB input gives 0dB output at zero phase.

5.4 Ideal Predistorter

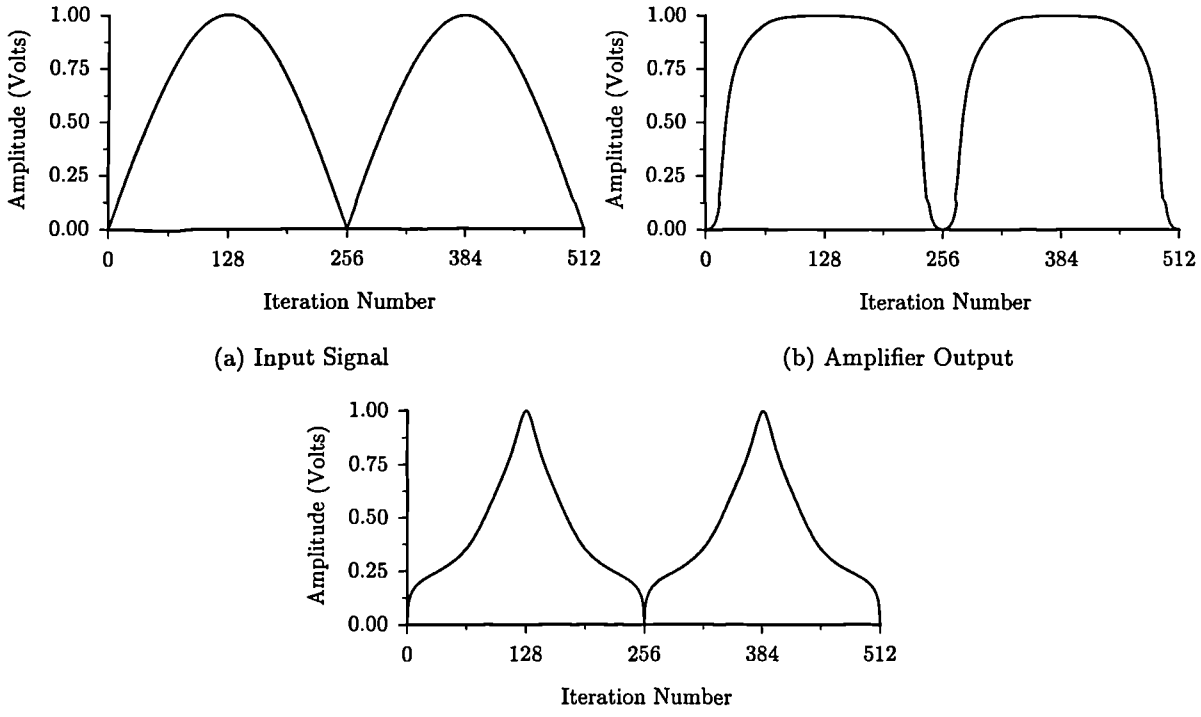
Using the theory presented in Section 5.2.1, the ideal predistorter transfer characteristic can be determined from the measured amplifier data. The measured transfer characteristic from the class C amplifier described in Section 4.3.3 is used, and is shown in quadrature form in Figure 5.5(a). The ideal predistorter characteristic derived from this is shown in Figure 5.5(b).



(a) Amplifier (b) Ideal Predistorter

Figure 5.5: Measured transfer characteristic of class C amplifier and the ideal predistorter transfer characteristic derived from it.

The envelopes of the two–tone input signal, the amplifier output and the predistorter output are shown in Figure 5.6. The amplifier output envelope shows a high degree of nonlinearity; the amplifier can be seen to have a cut–off region close to the zero–crossing points of the input envelope, and then a rapid change in output level which quickly results in saturation. This signal can be compared with that obtained from the practical feedforward system (Figure 4.14(a) on page 68). The general envelope shape is very similar, apart from the ripples which occur in the envelope of the practical amplifier output signal. These are probably due to a memory effect caused by saturation, which has not been accounted for in the (memoryless) amplifier model used in the simulation. The ideal predistorter envelope can be seen to compliment the gain expansion and compression evident in the amplifier envelope.

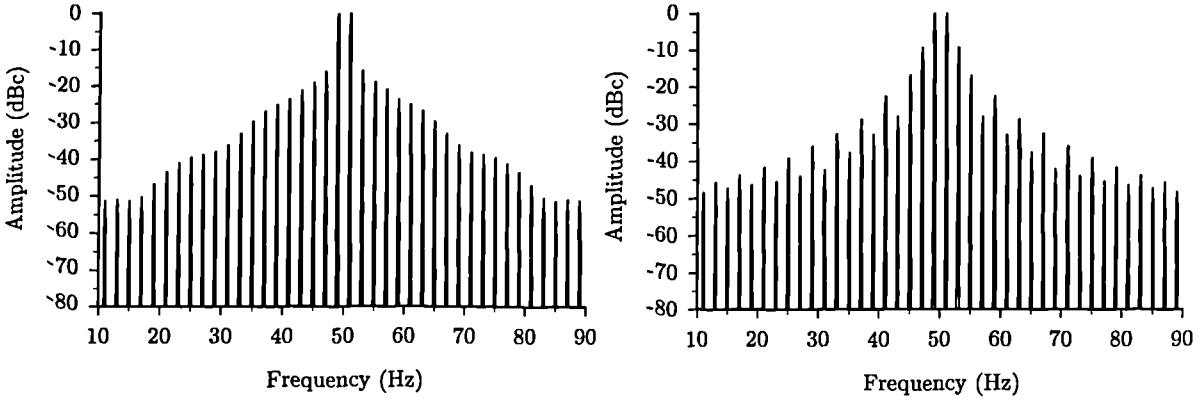


(a) Input Signal (b) Amplifier Output

(c) Predistorter Output

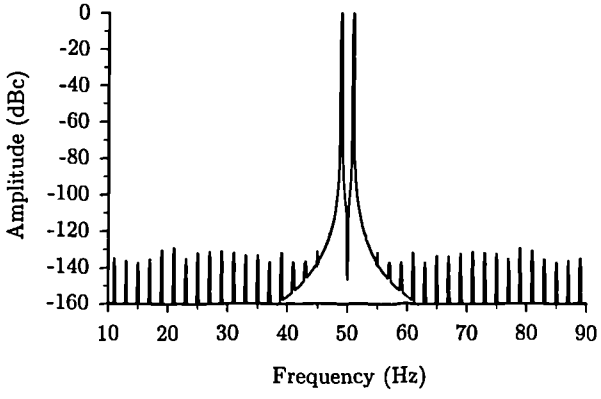
Figure 5.6: Envelope of signals.

When a two–tone input is applied to the amplifier and predistorter models, the spectra shown in Figure 5.7 result. The two input tones are at frequencies of 49Hz and 51Hz, and the FFT has been normalised such that these tones are at 0dB. Since the nonlinear models are memoryless, they exhibit no frequency–dependence, and thus the actual input frequencies are unimportant (the tone spacing is however an important consideration in the design of the digital filters required in the following chapter). The amplifier model produces third–order IMPs at −15.9dBc, which agrees well with the practical results. The output spectrum has a very similar form to that of the practical amplifier (Figure 4.12(a) on page 67). Thus the lack of modelling of the ripples present on the output envelope is unimportant, since their effect on the spectrum is negligible. The ideal predistorter output has third–order IMPs of −9.3dBc, and as would be expected, higher order IMPs are significant. The predistortion amplifier output spectrum using the ideal predistorter has IMPs below −120dBc. For an ideal predistorter it would be expected that no IMPs are produced at all. However, the amplifier characteristic is measured at a finite number of points, and the predistorter characteristic derived from these. Interpolation between the data points of both the amplifier and predistorter characteristics are required; this will introduce some error between them in interpolated regions, causing slight distortion. This is exacerbated by the different distributions of the abscissæ *e.g.*, if the amplifier input data points are equally spaced, the predistorter output data points will be equi–spaced due to the function inversion. The error introduced is negligible since the IMPs produced are below the −80dBc lowest output shown on all other FFT graphs. It would be expected that the error could be reduced by increasing the number of measured data points.



(a) Amplifier Output

(b) Predistorter Output



(c) Predistorted Amplifier Output

Figure 5.7: FFT of signals.

5.5 Efficiency Considerations of Predistortion Linearisation

The effect of predistortion on the efficiency of an amplifier is a complex issue. Both the instantaneous and average efficiencies are affected by predistortion, with the average efficiency being the most important factor to consider.

5.5.1 Predicting the Instantaneous Efficiency of a Predistortion Amplifier.

The instantaneous efficiency of a traditional amplifier class (A, AB, B or C), η_{INST}, is dependent upon the signal amplitude. In a nonlinear amplifier it is necessary to reference the instantaneous efficiency to either the input or the output amplitude, since the efficiency characteristic is modified by the nonlinear transfer characteristic. Thus, the instantaneous efficiency of an amplifier referenced to its input, η_{INST_i}, is given by

$$\eta_{INST_i} = \mathcal{E}_i(|w(t)|) \tag{5.20}$$

Whilst, the instantaneous efficiency of an amplifier referenced to its output, η_{INST_o}, is

$$\eta_{INST_o} = \mathcal{E}_o(|y(t)|) \tag{5.21}$$

This categorisation is not necessary in a linear amplifier since the two definitions of instantaneous efficiency are identical. The measured instantaneous efficiencies of the practical class A and C amplifiers previously described are shown in Figures 5.8 and 5.9 respectively.

Since the output of the predistortion amplifier is the same as that of the amplifier alone, the instantaneous efficiency of a predistortion amplifier referenced to its output is identical to that of the amplifier[1]. However, the instantaneous efficiency of a predistortion amplifier when referenced to its input is given by

$$\eta_{INST_i} = \mathcal{E}_i(\mathcal{F}_P(|x(t)|)) \tag{5.22}$$

In an ideal case, (*i.e.*, $\mathcal{F}_P = {\mathcal{F}_A}^{-1}$), the predistortion amplifier will be completely linear, and thus the instantaneous efficiency characteristic will be the same whether referenced to input or output.

5.5.2 Predicting the Average Efficiency of a Predistortion Amplifier.

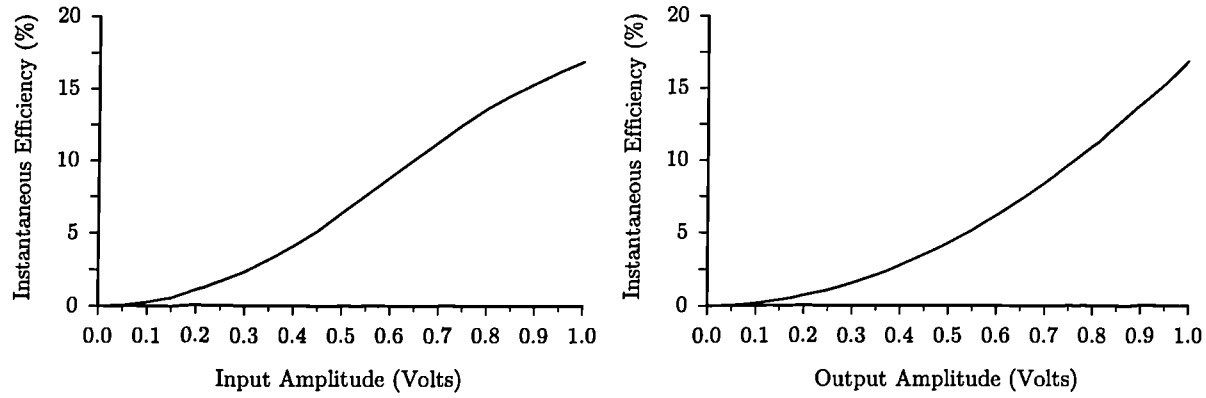
The average efficiency of an amplifier is defined as the ratio of the average RF output power to the average dc input power (Section 3.2.1), thus

$$\eta_{AVG} = \frac{P_{o,AVG}}{P_{dc,AVG}} \tag{5.23}$$

The instantaneous efficiency of an RF amplifier is dependent upon the instantaneous input envelope, and generally reaches peak efficiency when the amplifier is operating at its PEP. Therefore, the average efficiency is dependent not only on the instantaneous efficiency characteristic of the amplifier, but also upon the statistics of the input signal.

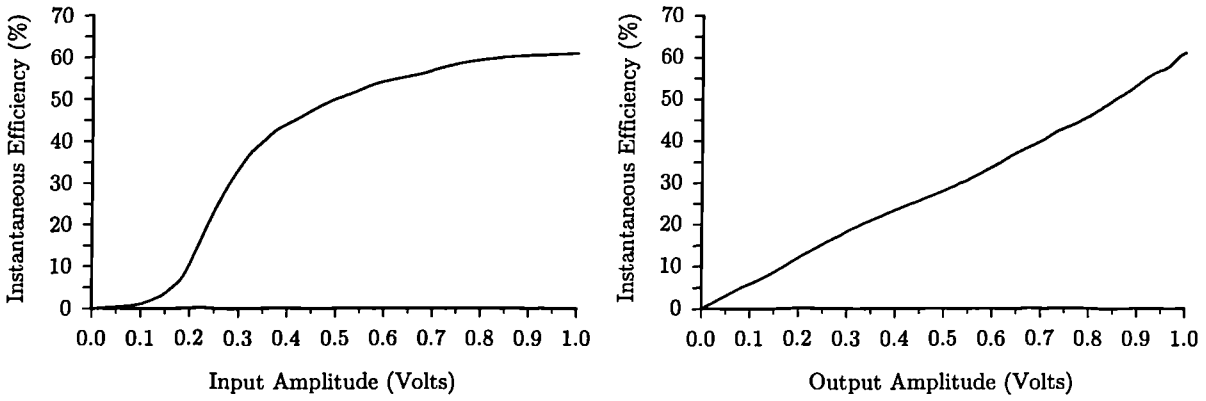
To investigate the statistics of the input and output signals of an amplifier, use may be made of *probability density functions* (p.d.f.s) [8]. The probability that the amplitude, A, of a signal

[1]This assumes that the predistortion network dissipates negligible power compared to the power supplied to the amplifier, which will generally be the case.



(a) Referenced to input (b) Referenced to output

Figure 5.8: Measured instantaneous efficiency of a class A amplifier.



(a) Referenced to input (b) Referenced to output

Figure 5.9: Measured instantaneous efficiency of a class C amplifier.

is within a certain range of amplitudes (A_1 to A_2) is given by

$$P(A_1 < A < A_2) = \int_{A_1}^{A_2} \rho(A)\, dA \tag{5.24}$$

where $\rho(\cdot)$ is the p.d.f. of the signal.

The average dc power supplied to the amplifier when driven by a signal with a p.d.f. of ρ_i is given by

$$P_{dc,AVG} = \int_0^{A_0} P_{dc}(A)\rho_i(A)\, dA \tag{5.25}$$

where $P_{dc}(A)$ is the instantaneous dc power supplied to the amplifier at amplitude A, and A_0 is the maximum amplitude of the input signal. Similarly, the average RF output power is given by

$$P_{o,AVG} = \int_0^{A_0} P_o(A)\rho_i(A)\, dA \tag{5.26}$$

where $P_o(A)$ is the instantaneous output power at an input amplitude A. Thus, the average efficiency is given by

$$\eta_{AVG} = \frac{\int_0^{A_0} P_o(A)\rho_i(A)\,dA}{\int_0^{A_0} P_{dc}(A)\rho_i(A)\,dA} \tag{5.27}$$

The instantaneous dc power supplied to the amplifier, P_{dc}, is given by

$$P_{dc}(A) = \frac{P_o(A)}{\eta_{INSTi}(A)} \tag{5.28}$$

where $\eta_{INSTi}(A)$ is the instantaneous efficiency of the amplifier operating at an input amplitude A.

The peak–to–mean ratio (ξ) of a signal may be found from [8]

$$\xi = \frac{1}{\int_0^1 A^2\rho(A)\,dA} \tag{5.29}$$

This provides a useful guide to the general statistics of the signal, and, it will be shown, also provides a guide to the average efficiency which can be attained.

Representative input signal p.d.f.s

To investigate the average efficiency of a predistortion amplifier, a number of representative p.d.f.s will be used, to simulate the effects of amplifying practical signals.

Ramp signal

All amplitudes of a ramp signal are equally likely, and thus the p.d.f. is given by

$$\rho_R(A) = 1 \tag{5.30}$$

This signal, whilst not being representative of a typical practical input signal, is useful in highlighting the nonlinear nature of the amplifier. The peak–to–mean ratio of a ramp signal is 4.8dB.

Two–tone signal

The p.d.f. of a two–tone signal is derived in [8], and is given by

$$\rho_{TT}(A) = \frac{2}{\pi}\frac{1}{\sqrt{1-A^2}} \tag{5.31}$$

This is one of the simplest non–constant envelope signals to generate practically, and since the envelope varies over its entire range it could provide a useful measure of the average efficiency. The peak–to–mean ratio of a two–tone signal is 3dB.

$(\pi/4)$–DQPSK signal

One of the most popular filtered digital modulation formats is $(\pi/4)$–Differential Quadrature Phase Shift Keying ($(\pi/4)$–DQPSK). Due to its filtered nature this signal is not constant–envelope, and thus has a non–singular p.d.f.. The p.d.f. of a filtered $(\pi/4)$–DQPSK signal (ρ_{QPSK}) may not be algebraically determined. Therefore, computer simulation of a $(\pi/4)$–DQPSK signal with random data has been performed using a root raised cosine filter with roll–off factors (α) of 0.35 and 0.5[2]. These are the roll–off factors used in practical systems

[2]The author gratefully acknowledges the assistance of Dr. Majid Boolorian for provision of the code to generate the $(\pi/4)$–DQPSK signals.

($\alpha = 0.35$ is used in the Trans-European Trunked RAdio (TETRA) [9] and Japanese Digital Cellular (JDC) [10] systems, whilst $\alpha = 0.5$ is used in the North American Digital Cellular (NADC) [11] system). The peak-to-mean ratio of the $(\pi/4)$-DQPSK signals has been found by simulation to be 3.04dB and 2.87dB for roll-off factors of 0.35 and 0.5 respectively.

Rayleigh distributed signal

The p.d.f. of a Rayleigh distributed envelope [8] is given by

$$\rho_{RY}(A) = 2A\xi \exp\left(-A^2\xi\right) \tag{5.32}$$

where ξ is the peak-to-mean ratio of the distribution in dB. The Rayleigh distribution approximates the p.d.f. of multi-carrier signals. The peak-to-mean ratio of a multi-carrier signal is simply equal to N, where N is the number of carriers. However, the Rayleigh p.d.f. extends to infinity, whereas in practice the probability of an amplitude greater than unity is zero, since the maximum amplitude is finite (in fact it is equal to N^2). This effect reduces the accuracy of the approximation at low peak-to-mean ratios, since the probability of an amplitude being greater than unity becomes more significant (4.2% with $\xi = 5$dB, falling to only 0.005% with $\xi = 10$dB).

These p.d.f.s are shown in Figure 5.10.

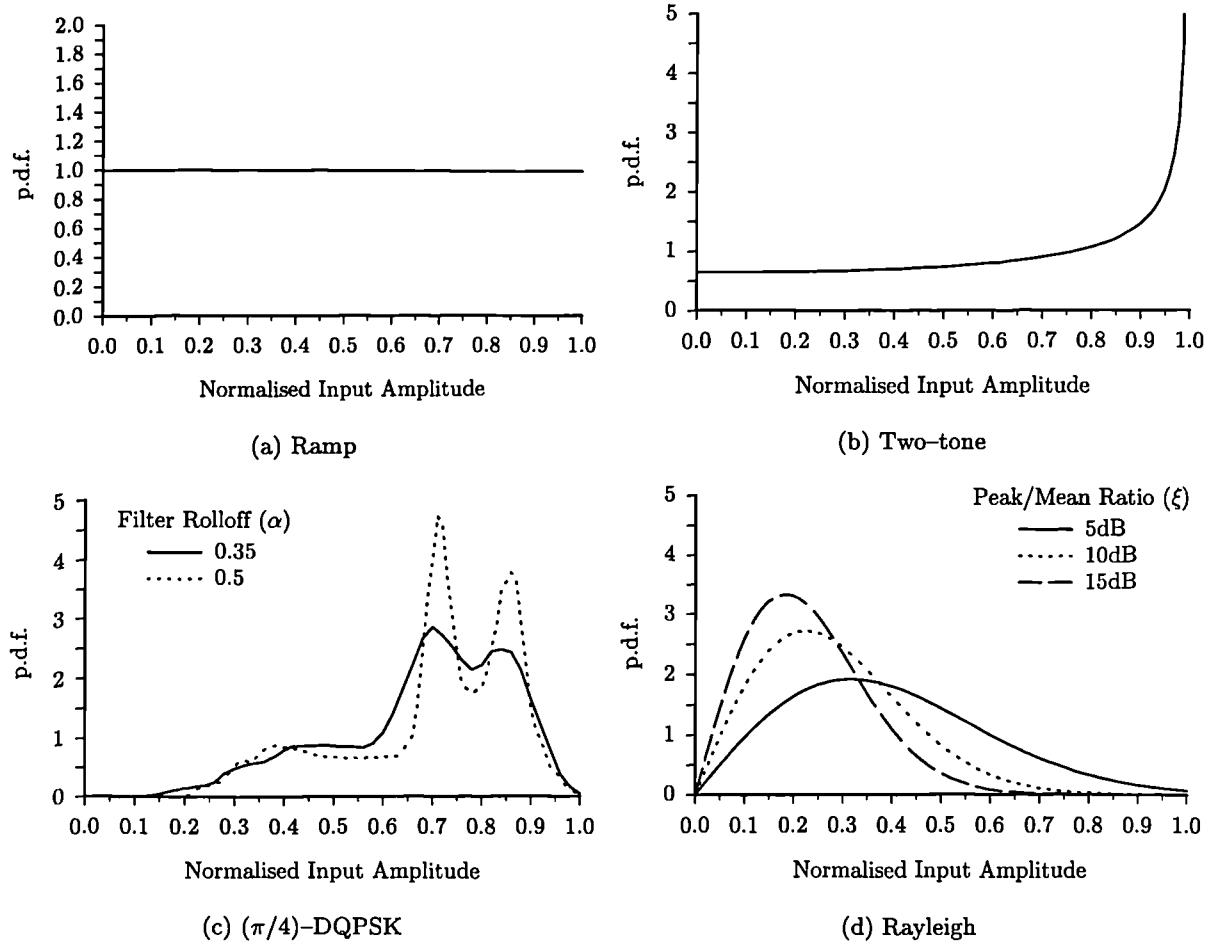


Figure 5.10: Probability density functions for the input signals.

Amplified signal p.d.f.s

The output signal of the amplifier will have a different p.d.f. than the input signal, due to the nonlinear nature of the amplifier. If a signal with a p.d.f. $\rho_i(x)$ is applied to a nonlinear function $y = \mathcal{F}(x)$, the p.d.f. of the output signal $\rho_o(y)$ can be shown to be given by [12]

$$\rho_o(y) = \frac{\rho_i\big(\mathcal{F}^{-1}(y)\big)}{\mathcal{F}'\big(\mathcal{F}^{-1}(y)\big)} \tag{5.33}$$

When the amplifier is backed–off from its saturated output, the calculation of the output p.d.f. is more complicated, and is derived below. In this case, the input p.d.f. remains the same, but the operating point on the nonlinear function is altered. To allow meaningful comparison the output p.d.f. is still defined from zero to one. Thus the nonlinear function, $\mathcal{F}(x)$ is rescaled to produce a new nonlinear function, $\mathcal{G}(x)$, given by

$$\mathcal{G}(x) = \frac{\mathcal{F}(A_0 x)}{\mathcal{F}(A_0)} \tag{5.34}$$

Thus, the derivative of this function is

$$\mathcal{G}'(x) = \frac{A_0 \mathcal{F}'(A_0 x)}{\mathcal{F}(A_0)} \tag{5.35}$$

and its inverse is given by

$$\mathcal{G}^{-1}(y) = \frac{\mathcal{F}^{-1}(y\mathcal{F}(A_0))}{A_0} \tag{5.36}$$

Therefore

$$\mathcal{G}'\big(\mathcal{G}^{-1}(y)\big) = \frac{A_0 \mathcal{F}'\big(A_0 \mathcal{G}^{-1}(y)\big)}{\mathcal{F}(A_0)} \tag{5.37}$$

By relevant substitution and simplification this becomes

$$\mathcal{G}'\big(\mathcal{G}^{-1}(y)\big) = \frac{A_0 \mathcal{F}'\big(\mathcal{F}^{-1}(y\mathcal{F}(A_0))\big)}{\mathcal{F}(A_0)} \tag{5.38}$$

Thus, the output p.d.f. is given by

$$\rho_o(y, A_0) = \frac{\mathcal{F}(A_0)\rho_i\Big(\frac{\mathcal{F}^{-1}(y\mathcal{F}(A_0))}{A_0}\Big)}{A_0 \mathcal{F}'\big(\mathcal{F}^{-1}(y\mathcal{F}(A_0))\big)} \tag{5.39}$$

When the representative input signals are amplified by the class C amplifier model introduced in Section 5.3.2, the output signal p.d.f.s are found to be radically different from the input signal p.d.f.s, as depicted in Figure 5.11 for the ramp and two–tone signals (the $(\pi/4)$–DQPSK and Rayleigh amplifier output p.d.f.s are shown in Appendix C). The ramp signal output allows visualisation of the effect of the amplifier nonlinearity, without being biased by the input p.d.f.. The amplifier nonlinearity can be seen to increase the probability of the output amplitude being near its maximum or zero, due its switch–like transfer characteristic. At such extreme amplitudes the amplifier will be operating more efficiently than at intermediate levels. Therefore, it is clear that the action of predistortion will reduce the average efficiency of the amplifier, since the amplifier will no longer be operating at these extremes for the majority of the time.

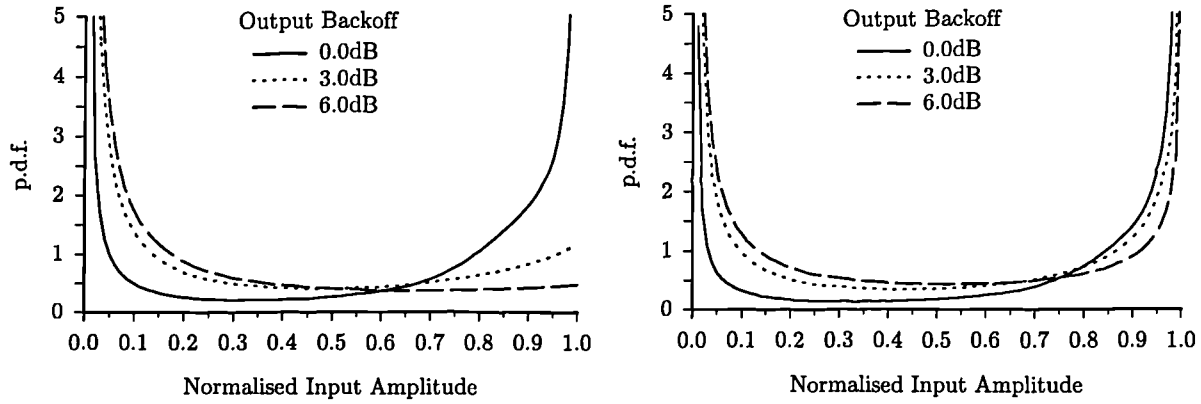


Figure 5.11: Class C amplifier output p.d.f.s for the ramp and two–tone input signal p.d.f.s.

The output p.d.f. is dependent upon the output back–off applied to the amplifier; this is not the case with the input p.d.f., nor with the ideal predistortion amplifier. As the amplifier is backed–off the probability of amplitudes near maximum reduces, since the amplifier is not saturating to the same degree.

The ideal predistorter p.d.f.s for the ramp and two–tone input signals are shown in Figure 5.12 (the $(\pi/4)$–DQPSK and Rayleigh amplifier output p.d.f.s are shown in Appendix C). The output back–off quoted for the graphs is that of the predistortion amplifier. This will obviously be a different input back–off than for the unpredistorted case shown above. It can be seen that to compensate for the amplifier nonlinearity increasing the p.d.f. at the maximum and minimum amplitudes, the predistorter enhances the probability of signals in the middle of the range. This is easiest to discern on the ramp p.d.f. graph depicted in Figure 5.12(a). The predistorter can be seen to shift the p.d.f. to higher amplitudes when the amplifier is backed–off; this compensates for the amplifier causing a shift of the p.d.f. to lower amplitudes.

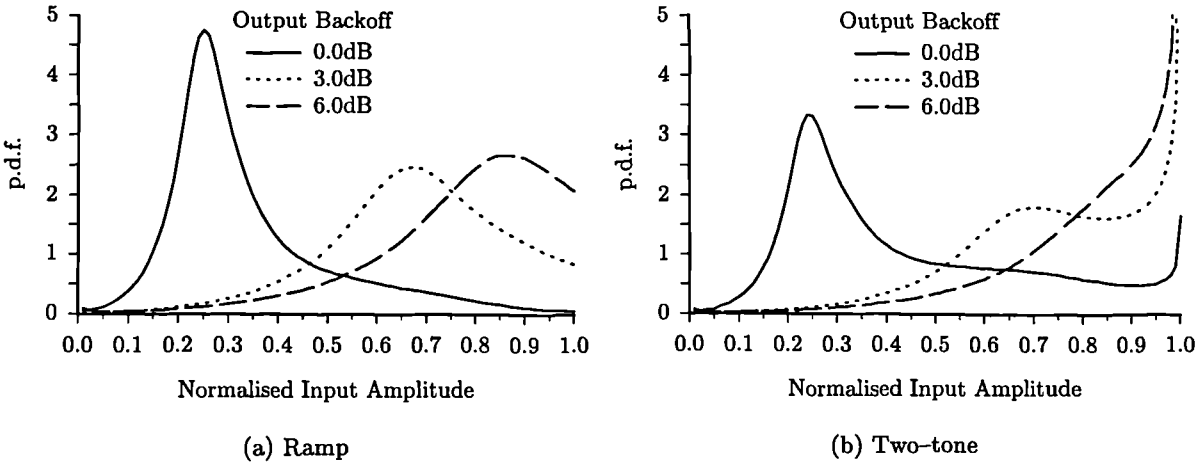


Figure 5.12: Predistorter output p.d.f.s for the ramp and two–tone input signal p.d.f.s.

5.5.3 Calculated average efficiency results

The average efficiency of the class A and C amplifiers with and without predistortion are calculated and shown in Figure 5.13. The maximum average efficiency of the predistortion amplifier is lower than that of the amplifier alone. As the amplifier is backed–off, the average efficiency of the amplifier falls more quickly than that of the predistortion amplifier. However, care should be taken interpreting this, since the input back–off levels will be different in the two cases.

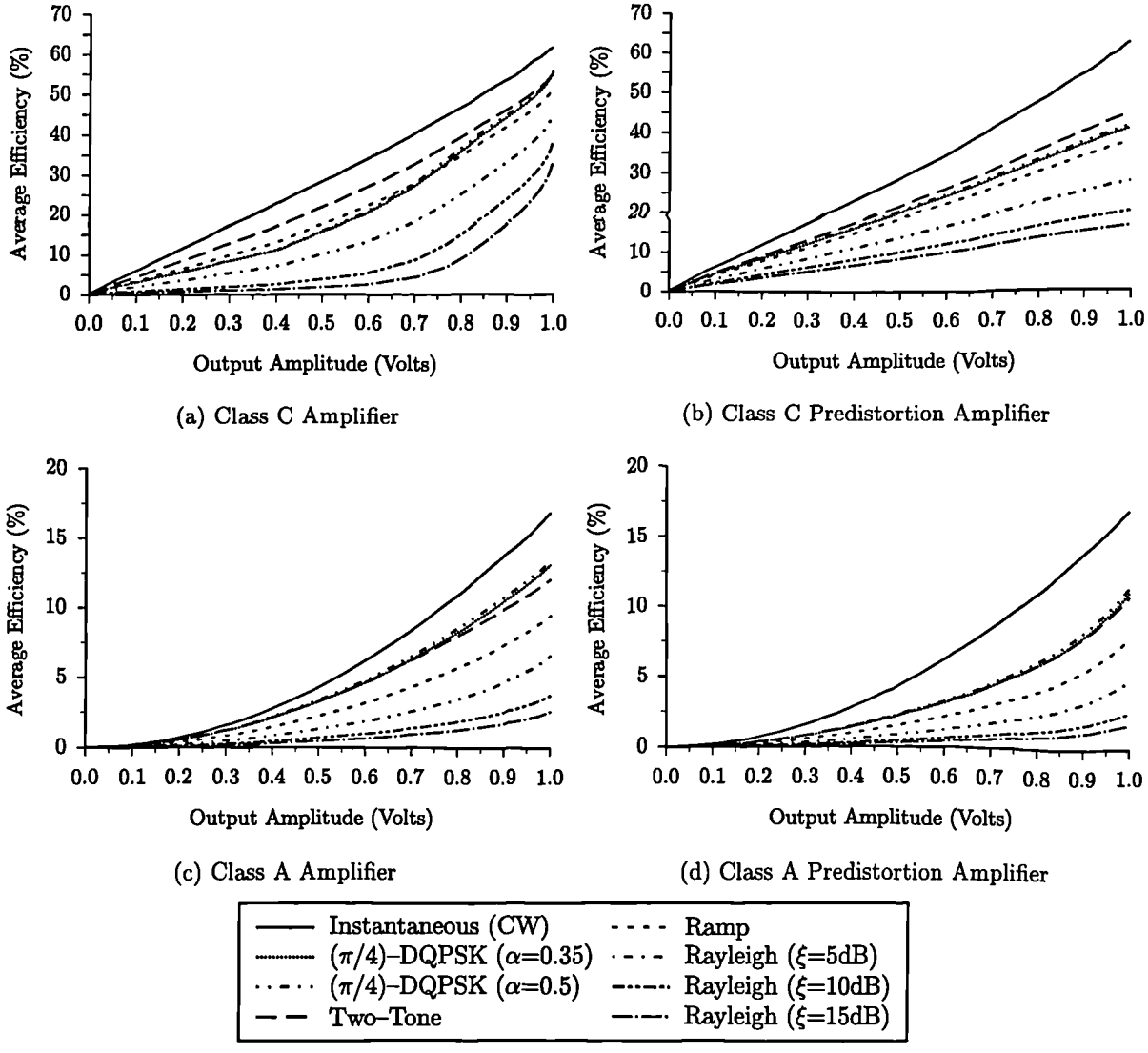


Figure 5.13: Calculated average efficiency of class C and class A amplifiers with and without predistortion.

The two–tone and $(\pi/4)$–DQPSK signals produce the highest average efficiency, since the signals spend most of the time near to the maximum amplitude (*i.e.*, low peak–to–mean ratios), where the instantaneous efficiency is highest.

It is interesting to note that the difference in average efficiency using $(\pi/4)$–DQPSK with either filter roll–off factor is less than 1% and thus insignificant. Consequently, the average efficiency does not affect the choice of either roll–off factor in a system specification.

In a multi–carrier system the probability of getting high amplitudes reduces markedly as the

number of carriers is increased. Consequently it may be acceptable to limit the input signal to the amplifier, reducing the peak–to–mean ratio and so increasing average efficiency. In the approximate multi–carrier case, using the Rayleigh distribution, limiting the signal is effectively the same as using a Rayleigh distribution with a reduced peak–to–mean ratio. To illustrate the effect of limiting the input signal in a true multi–carrier system on the average efficiency and linearity would require considerable further simulation study. The effect of limiting on the linearity has in fact been examined in [13]; however, no efficiency simulation has yet been presented.

To ascertain more clearly the effect of predistortion of an amplifier on its average efficiency, it is useful to consider the ratio of the average efficiency of the predistortion amplifier to that of the amplifier alone. This allows the effect of peak–to–mean ratio, and amplifier nonlinearity to be more clearly examined. Figure 5.14 illustrates that as the peak–to–mean ratio increases, predistortion causes increased degradation of the average efficiency. It may also been seen that there is less reduction in average efficiency when applying predistortion to a reasonably linear amplifier. This is to be expected, since there is less difference in transfer characteristics between the amplifier and predistortion amplifier.

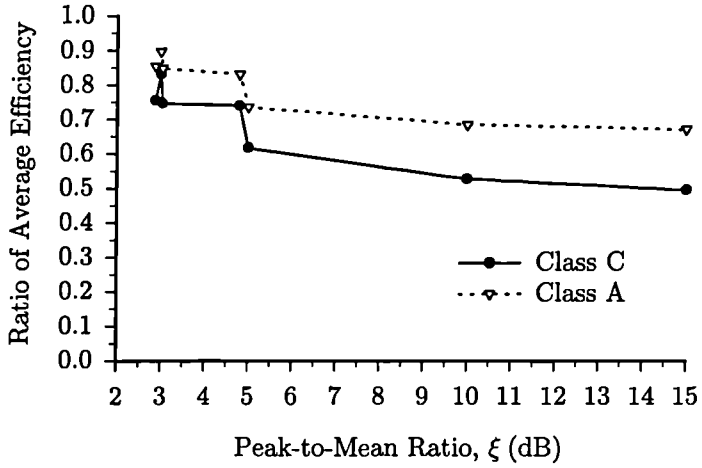


Figure 5.14: Ratio of the average efficiency of the predistortion amplifier to that of the amplifier alone as a function of the peak–to–mean ratio of the input signal, at peak output.

The comparisons made above assume that the amplifier and predistortion amplifier are being operated at the same output back–off level. However, if a certain linearity performance is required, it is possible to operate the predistortion at a lower output back–off level than the amplifier alone. This will offset the degradation in average efficiency caused by applying predistortion, and may in some situations allow the predistortion amplifier to operate more efficiently than the amplifier alone.

The main amplifier in the practical feedforward system described in Section 4.3.3, had an average efficiency of 55% when being driven by a two–tone input signal; this compares favourably with the calculated average efficiency, depicted in Figure 5.13(a). The theoretical average efficiency of the feedforward amplifier is 31%; thus the ratio of average efficiency of the feedforward amplifier to the average efficiency of the main amplifier is 0.56. In the case of applying predistortion, the ratio is 0.88. Therefore, the reduction in average efficiency caused by applying linearisation is much less for predistortion than for feedforward linearisation in the class C amplifier system.

5.6 Summary

This chapter has investigated the various nonlinear elements and architectures which can be used to implement a predistortion system. It is evident that the best linearity performance could be achieved with the use of a generic vector predistorter. This would allow significant reduction of both AM/AM and AM/PM conversion effects. The networks currently used in generic predistortion have been shown to be inflexible and provide limited linearity improvement. The following chapter will describe an alternative generic predistortion network which offers the potential of improved linearity and flexibility over current methods.

The transfer characteristic of an ideal predistorter has been derived, and a computer simulation of an ideal predistortion amplifier using a measured class C amplifier *transfer characteristic* has been presented. It has been shown that, due to the characteristics of the approximating function, the simulated 'ideal' predistorter, still generates IMPs; however, the IMP level is so low as to be negligible.

The p.d.f.s of various representative input signals have been used to illustrate the effects of the nonlinearity of both the amplifier and the ideal predistorter. These have been used to allow prediction of the average efficiency of an amplifier with and without predistortion being applied.

The peak–to–mean ratio of the input signal gives an indication of the average efficiency which can be achieved. A low peak–to–mean ratio implies that the amplifier operates close to its maximum amplitude (and hence maximum efficiency) for the majority of the time, and so has a relatively high average efficiency.

Finally, a comparison of the reduction in average efficiency caused by applying either predistortion or feedforward linearisation has been presented. It has been found that feedforward has a far greater impact on the average efficiency than is the case with predistortion, and therefore, predistortion would be the preferred technique for high efficiency applications, assuming that satisfactory linearity performance can be obtained.

REFERENCES

[1] M. KUMAR, J.C. WHARTENBY, AND H.J. WOLKSTEIN. Predistortion linearizer using GaAs dual-gate MESFET for TWTA and SSPA use in satellite transponders. *IEEE Transactions on Microwave Theory and Techniques*, MTT-33(12):1479–1488, December 1985.

[2] A. EGGER, M. HORN, AND T. VIEN. Broadband linearisation of microwave power amplifiers. In *10th European Microwave Conference*, pages 490–494, Warsaw, Poland, September 1980.

[3] S. AIHARA, T. NISHIUMI, Y. FUJIKI, AND S. FUKUDA. GaAs FET power amplifiers as substitutes for TWT amplifiers in a multi-level QAM digital radio system. In *IEEE International Conference on Communications*, volume 1, pages 1.2.1–1.2.5, Seattle, Washington, June 1987.

[4] G.M. BLAIR, J.J. DALY, AND J.F. MOSS. Design and characterisation of a GaAs FET power amplifier for a 64 QAM digital radio transmitter. In *IEEE International Conference on Communications*, volume 1, pages 1.4.1–1.4.6, Seattle, Washington, June 1987.

[5] HIDEO ASHIDA, YOSHIAKI SUZUKI, ISAMU UMINO, AND NORIO TOZAWA. C-Band 100 Watt GaAs FET amplifier for digital radio. In *IEEE International Conference on Communications*, pages 1/23–1/27, Seattle, Washington, June 1987.

[6] J. NAMIKI. An automatically controlled predistorter for multilevel quadrature amplitude modulation. *IEEE Transactions on Communications*, COM-31(5):707–712, May 1983.

[7] T. NOJIMA AND T. KONNO. Cuber predistorter linearizer for relay equipment in 800 MHz band land mobile telephone system. *IEEE Transactions on Vehicular Technology*, VT-34(4):169–177, November 1985.

[8] F.H. RAAB. Average efficiency of power amplifiers. In *Proc. of RF Technology Expo 86*, pages 473–486, Anaheim, CA, USA, February 1986. Cardiff Publishing Co.

[9] *ETSI Draft Specification for Trans-European Trunked Radio.* ETS 300 392-2:1994.

[10] *RCR STD-27A*, January 1992.

[11] *EIA/TIA Interim Standard*, 1994. IS54-B.

[12] H. STARK AND J.W. WOODS. *Probability, random processes, and estimation theory for engineers.* Prentice-Hall, New Jersey, USA, 1986.

[13] D.W. BENNETT. *The design and performance of automatically-controlled feedforward amplifiers.* PhD thesis, Dept. of Electrical & Electronic Engineering, University of Bristol, September 1995.

Chapter Six

Piecewise–Linear Predistortion

This chapter presents an alternative analogue predistortion method, known as *piecewise-linear predistortion*, which offers the potential for improved performance compared with previous methods. This technique is described, and possible architectures for its practical implementation are suggested. Computer simulation is used to illustrate the linearity improvement which the technique offers. A number of different adaption schemes which allow optimisation of the predistortion amplifier linearity are described.

6.1 Introduction

The previous chapter has described the forms of predistortion network in use at present. To achieve optimal linearity, a vector predistorter should be used, although in the literature the scalar predistorter is prevalent, due to its simplicity and cost. The predistorter network nonlinearity also has two basic forms; the polynomial and generic types. The polynomial predistorter uses individual orders of nonlinearity (usually only third), and can only affect the distortion levels of that order or less. Therefore, the reduction of multiple orders of distortion requires the use of a number of nonlinearities, each of which requires separate control. As a consequence, the technique is impractical for reducing high numbers of orders of distortion, such as that present in class C amplifiers. The generic predistorter uses a nonlinear device whose characteristics can be altered, *e.g.*, by changing the bias point of the device, with the device characteristic being chosen to be an approximate inverse to that of the amplifier. However, since this is only an approximation, it is possible for some IMP orders to actually increase, whilst others are reduced. Therefore, it is necessary to compromise to achieve good performance. The major problem with contemporary generic predistorter networks is that their control mechanism is crude, with only one or two degrees of freedom. This does not generally allow the inverse characteristic to be well approximated, with obvious implications on linearity.

An improved generic predistorter has been suggested in [1], and further developed in [2], which increases the number of degrees of freedom, providing improved flexibility and linearity, at the expense of increased complexity. The technique uses a nonlinear network, termed a *piecewise–linear predistorter* (PLP), which approximates the ideal predistorter characteristic using a number of linear segments. The method is completely general and so can be used to linearise any kind of memoryless nonlinearity. This is not the case with standard generic predistorters which tend to be designed for a specific nonlinearity.

This technique has not been analysed in any detail in the literature, and so the potential linearity improvement which the technique may provide is not known. This chapter describes the PLP method, and derives a number of alternative configurations to that previously published. All of these PLP configurations are simulated to predict their linearity improvement, using a number of optimisation schemes.

6.2 Piecewise–Linear Predistortion

The piecewise–linear predistortion technique approximates the ideal predistorter characteristic using a number of piecewise–linear *segments*, with the transition between segments occurring at a *knot* [3]. The characteristic of a general PLP element is a discontinuous function and is shown in Figure 6.1. Since the segments are piecewise–linear this is termed a *discontinuous first–order* PLP.

The general equation for a PLP with an input amplitude A in a given segment, n, is simply that for linear interpolation, and is given by

$$\mathcal{F}(A) = \mathcal{F}(A_L[n]) + \left(\frac{A - A_L[n]}{A_U[n] - A_L[n]}\right)\left(\mathcal{F}(A_U[n]) - \mathcal{F}(A_L[n])\right) \tag{6.1}$$

where $A_L[n]$ and $A_U[n]$ are the abscissæ of the lower and upper knots bounding the input A respectively. If the transition between knots n and $n+1$ occurs at an amplitude $A[n]$, then

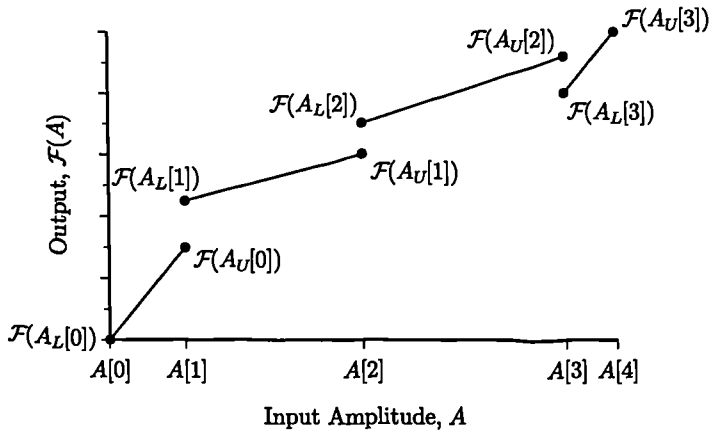


Figure 6.1: Discontinuous first–order PLP characteristic.

it is necessary that

$$A_L[n] = A[n]^+$$
$$A_U[n] = A[n+1]^-$$

The value of the function at the transition, $A[n]$, may be undefined, if the PLP function is discontinuous. Thus, Equation 6.1 simplifies to

$$\mathcal{F}(A) = \mathcal{F}(A_L[n]) + \left(\frac{A - A[n]}{A[n+1] - A[n]}\right)\left(\mathcal{F}(A_U[n]) - \mathcal{F}(A_L[n])\right) \tag{6.2}$$

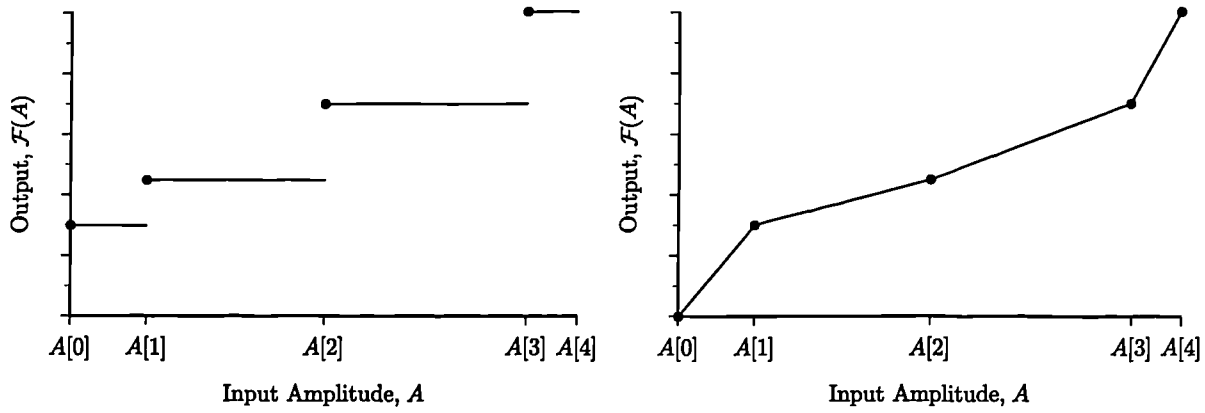
The PLP characteristic is normalised such that the range of the function is $[0:1]$. To define an N–segment PLP requires a set of N equations, given by

$$\mathcal{F}(A) = \begin{cases} \mathcal{F}(0) + \left(\frac{A}{A[1]}\right)\left(\mathcal{F}(A_U[0]) - \mathcal{F}(0)\right) & 0 \le A < A[1] \\ \mathcal{F}(A_L[1]) + \left(\frac{A-A[1]}{A[2]-A[1]}\right)\left(\mathcal{F}(A_U[1]) - \mathcal{F}(A_L[1])\right) & A[1] \le A < A[2] \\ \vdots & \\ \mathcal{F}(A_L[N-1]) + \left(\frac{A-A[N-1]}{1-A[N-1]}\right)\left(\mathcal{F}(1) - \mathcal{F}(A_L[N-1])\right) & A[N-1] \le A \le 1 \end{cases} \tag{6.3}$$

The general PLP characteristic is discontinuous, however, constraints may be added to derive other architectures which are simpler to implement practically. The simplest PLP uses piecewise–constant segments, and is termed a *zero–order* PLP. Another alternative is to constrain the characteristic to be continuous, forming the *continuous first–order PLP* . Example characteristics of these PLP configurations are depicted in Figure 6.2.

6.3 Practical PLP Architectures

Although only one practical circuit topology (the attenuator–based resistor–diode network) has been suggested in the literature to construct a PLP, there are a number of alternative circuit architectures which may be suitable. These alternative circuits have been designed for two main applications; analogue computing and triangular–to–sinewave conversion. All of the



(a) Zero–order (b) Continuous first–order

Figure 6.2: Alternative PLP characteristics.

circuits presented are for generation of a continuous first–order PLP, although modification may be possible to enable generation of the other PLP configurations. It should be noted that a discontinuous PLP will be more complex than a continuous or zero–order PLP since more knots are required to define the segments.

6.3.1 Resistor–Diode PLP

This PLP configuration is based on attenuator or amplifier circuits with the addition of one or more resistor–diode networks [4–7] as shown in Figure 6.3. The basis for operation is to bias the diodes at different voltages, so that they switch on at differing input voltages, hence altering the transfer characteristic of the circuit.

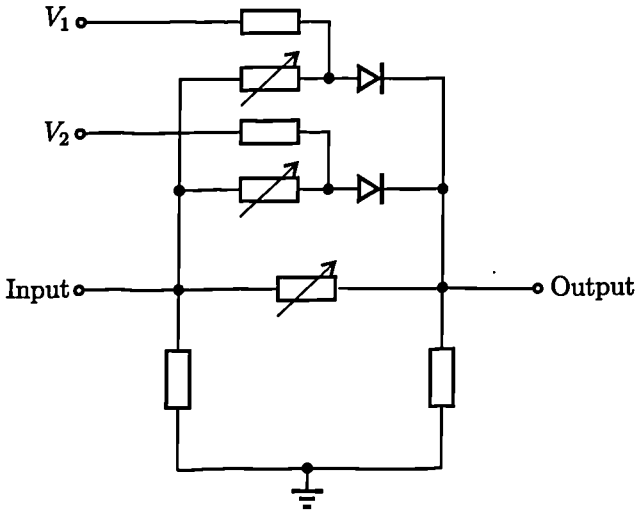
The circuits shown are only applicable for use with unipolar signals, *i.e.*, the envelope, but can be converted to operate with bipolar signals, *i.e.*, the instantaneous signal, by using an additional resistor–diode network to operate on negative input signals. These configurations are only suitable for generating monotonic–increasing functions. In order to produce a more general non–monotonic function, an additional resistor–diode network is required, in parallel with the input resistor to ground in the attenuator case, or the feedback resistor in the amplifier case. The transfer characteristic of the amplifier–based circuit is derived in Appendix D.

To improve the high–frequency performance of the circuits, Schottky diodes should be used, since these have switching speeds of the order of picoseconds [8]. The amplifier circuit has the obvious disadvantage that the bandwidth of operation will be affected by the performance of the operational amplifier. However, in recent years the unity gain bandwidth (UGBW) of op–amps has increased markedly, and currently devices with a UGBW greater than 1GHz can be obtained [9]. It would be expected that the UGBW will continue to rise steadily.

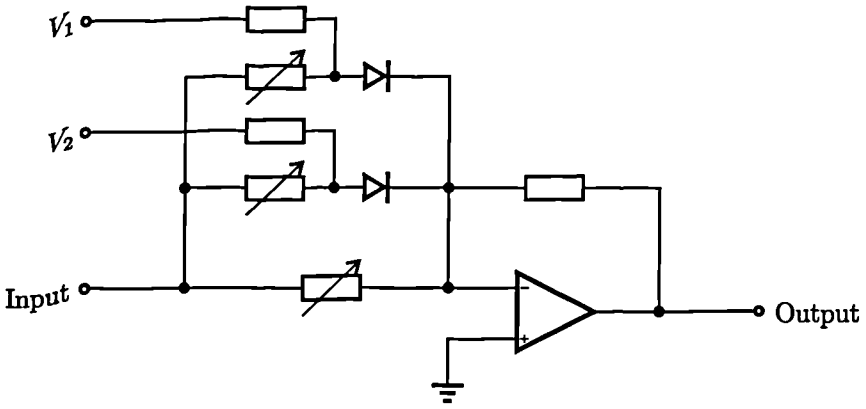
An alternative technique is to use zener diodes in the circuit, alleviating the need for bias voltages [10]. However, this would result in an inflexible circuit, whose speed of operation is limited.

6.3.2 Differential–Pair PLP

This configuration, depicted in Figure 6.4, uses a differential–pair and a resistor–diode network to produce a piecewise–linear characteristic [11]. The circuit requires both the input signal,



(a) Attenuator form.



(b) Amplifier form.

Figure 6.3: Two forms of resistor–diode PLPs.

and its antiphase equivalent to be applied to the differential–pair. The circuit shown allows for monotonic–decreasing functions to be generated.

The circuit operates essentially as a switched common–emitter amplifier, and can operate with bipolar signals. The antiphase equivalent of the input signal may be accurately generated over a broad frequency band by the use of a 180° hybrid [12].

6.3.3 Limiter PLP

This configuration, depicted in Figure 6.5, uses a number of limiter circuits, each limiting at different input and output voltages [13]. The outputs of the limiters are then summed, to effectively generate a piecewise–linear function. The circuit can operate with either unipolar or bipolar signals.

The circuit shown will generate a monotonic–decreasing function. In order to generate non–monotonic functions, an additional set of limiters would be required, whose outputs are subtracted from the original limiters' outputs.

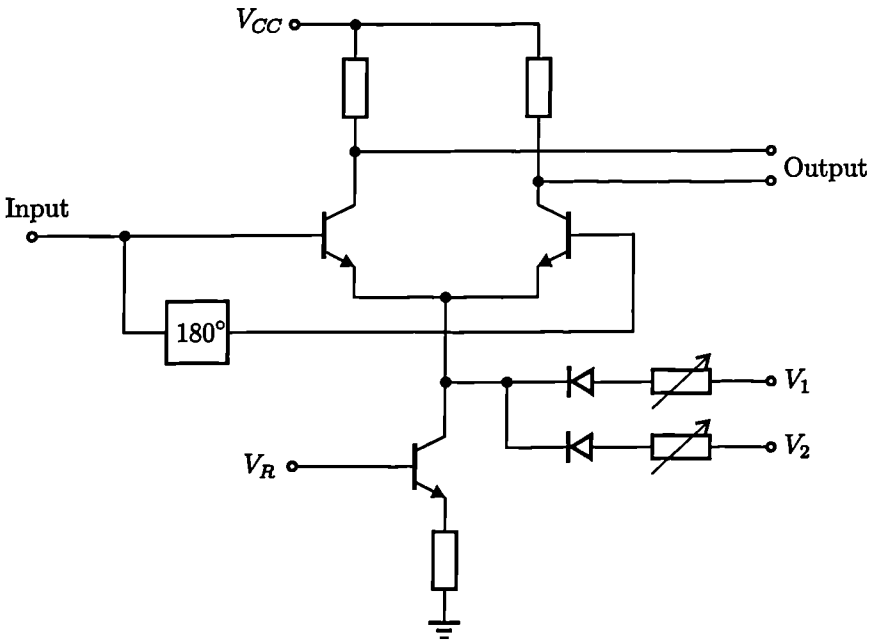


Figure 6.4: Configuration of a differential–pair PLP.

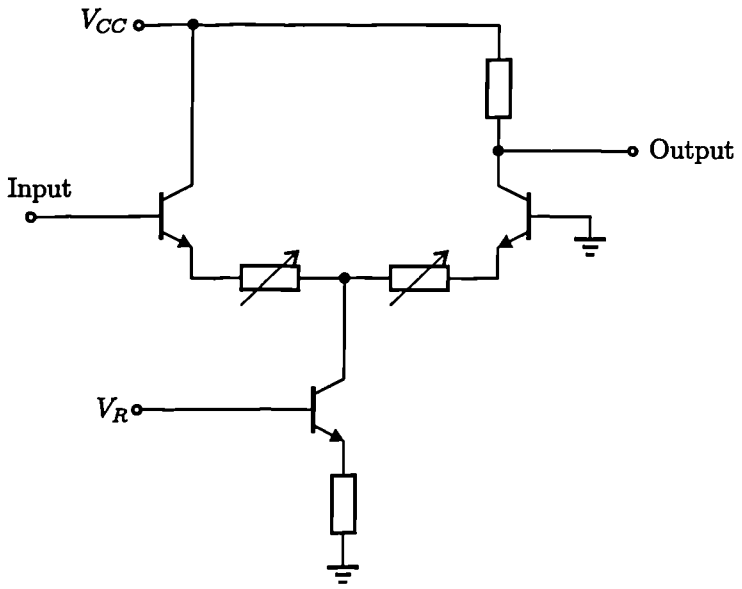


Figure 6.5: Basic building block for a limiter PLP.

6.3.4 Practical Implementation Issues

The PLP architectures described above are considerably more complex than a standard generic predistorter. However, they are completely general, and so can be used with any nonlinear transfer characteristic. Such is not the case with standard generic predistorters, which tend to be constructed for use with a particular nonlinearity. The circuits may be easily integrated, allowing a flexible monolithic predistorter to be produced.

6.4 Piecewise–Linear Predistorter Model

The predistorter model used is the complex envelope equivalent of the quadrature amplifier model derived in Section 2.4.3, since this has been shown to be the most appropriate method of achieving a practical vector predistorter (Section 5.2.3). It is possible to produce a piecewise–linear approximation to either the complex gain or the transfer characteristic of the ideal predistorter. This gives rise to two generic types of PLP, designated the gain–type PLP and transfer characteristic–type PLP respectively[1]. As with the amplifier model, the envelope of the input signal is used to determine the output of the nonlinear function blocks, and hence, the predistorter is a bandpass memoryless nonlinearity. The use of the envelope to determine the predistorter output will limit the bandwidth over which the practical predistorter will operate, since the envelope will occupy a wider band than the input signal.

6.4.1 Gain–Type PLP

The gain–type PLP (G–PLP) uses the envelope of the signal to determine the complex gain (G_P) of the predistorter block. The output signal, $y(t)$, for a given input signal, $x(t)$, is given by

$$y(t) = x(t)G_P(A) \tag{6.4}$$

where A is the envelope of the input signal, and is equal to $|x(t)|$. A practical implementation of the G–PLP is shown in Figure 6.6(a). The input signal envelope is determined using an envelope detector, and applied to two PLP elements, which are then mixed with quadrature versions of the input signal, and summed to produce the predistorter output signal. The complex envelope equivalent, which is to be used in the simulation, is shown in Figure 6.6(b).

The three possible configurations for the PLP elements in the G–PLP are described below.

Zero–order

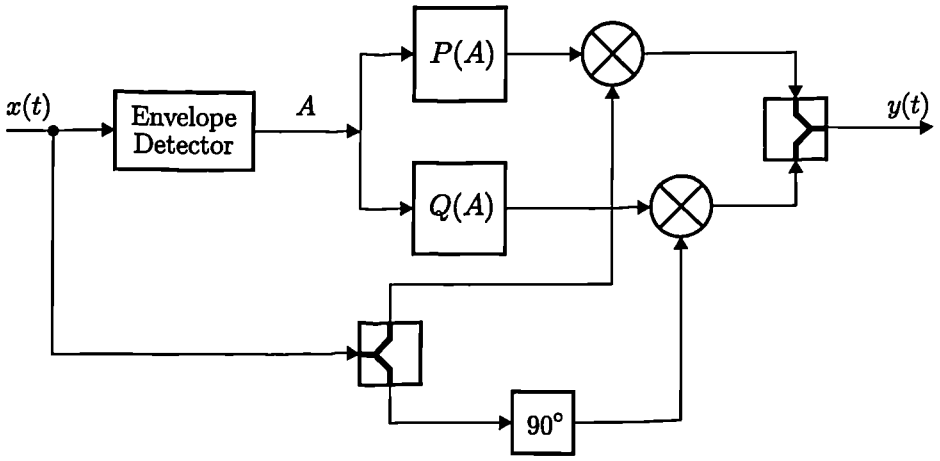
The zero–order G–PLP utilises constant complex gain within each segment. This is the type of predistorter which is used in adaptive baseband predistortion [14]. The complex gain of the predistorter is given by

$$G_P(A) = \begin{cases} G_P(0), & 0 \le A < A[1] \\ G_P(A[1]), & A[1] \le A < A[2] \\ \quad \vdots & \\ G_P(A[N-1]), & A[N-1] \le A \le 1 \end{cases} \tag{6.5}$$

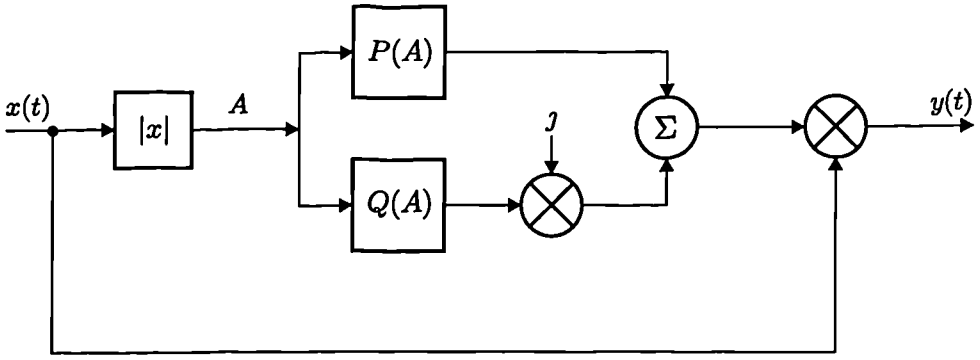
Continuous first–order

The continuous first–order G–PLP (henceforth simply the continuous G–PLP) uses a complex gain which changes linearly within each segment, but is constrained to be continuous at each

[1]It should be noted that the only PLP configuration presented in the literature is that of the continuous transfer characteristic–type PLP.



(a) Practical configuration



(b) Complex envelope equivalent form

Figure 6.6: Block diagram of the gain–type PLP (G–PLP).

knot transition. Thus, the complex gain is given by

$$G_P(A) = \begin{cases} G_P(0) + \left(\frac{A}{A[1]}\right)\left(G_P(A[1]) - G_P(0)\right), & 0 \le A < A[1] \\ G_P(A[1]) + \left(\frac{A-A[1]}{A[2]-A[1]}\right)\left(G_P(A[2]) - G_P(A[1])\right), & A[1] \le A < A[2] \\ \vdots & \\ G_P(A[N-1]) + \left(\frac{A-A[N-1]}{1-A[N-1]}\right)\left(G_P(1) - G_P(A[N-1])\right), & A[N-1] \le A \le 1 \end{cases} \tag{6.6}$$

This type of characteristic is easily generated using the analogue networks presented in the previous section.

Discontinuous first–order

The discontinuous first–order G–PLP (henceforth simply the discontinuous G–PLP) is similar to the continuous G–PLP, but does not have the restriction of maintaining a continuous characteristic, *i.e.*, the characteristic is only piecewise–continuous. Consequently, the predistorter

complex gain is given by

$$G_P(A) = \begin{cases} G_P(0) + \left(\frac{A}{A[1]}\right)(G_P(A_U[0]) - G_P(0)), & 0 \leq A < A[1] \\ G_P(A_L[1]) + \left(\frac{A-A[1]}{A[2]-A[1]}\right)(G_P(A_U[1]) - G_P(A_L[1])), & A[1] \leq A < A[2] \\ \vdots & \\ G_P(A_L[N-1]) + \left(\frac{A-A[N-1]}{1-A[N-1]}\right)(G_P(1) - G_P(A_L[N-1])), & A[N-1] \leq A \leq 1 \end{cases} \tag{6.7}$$

6.4.2 Transfer Characteristic–Type PLP

The transfer characteristic–type PLP (TC–PLP) uses the envelope of the input signal to determine the envelope and phase shift of its output. The output signal, $y(t)$, for a given input signal, $x(t)$, is given by

$$y(t) = \left(\frac{x(t)}{A}\right) T_P(A) \tag{6.8}$$

where $T_P(A)$ is the quadrature representation of the envelope transfer characteristic of the PLP. A practical implementation of the TC–PLP is shown in Figure 6.7(a). The input signal envelope is determined using an envelope detector, and applied to two PLP elements in the same manner as with the G–PLP. The PLP outputs are mixed with quadrature versions of the phase information of the input signal (determined using a limiter), and summed to produce the predistorter output signal. The complex envelope equivalent, which is to be used in the simulation, is shown in Figure 6.7(b). This configuration is more complex than the G–PLP in a practical system, due to its use of a limiter.

The three possible configurations for the PLP elements in the TC–PLP are described below.

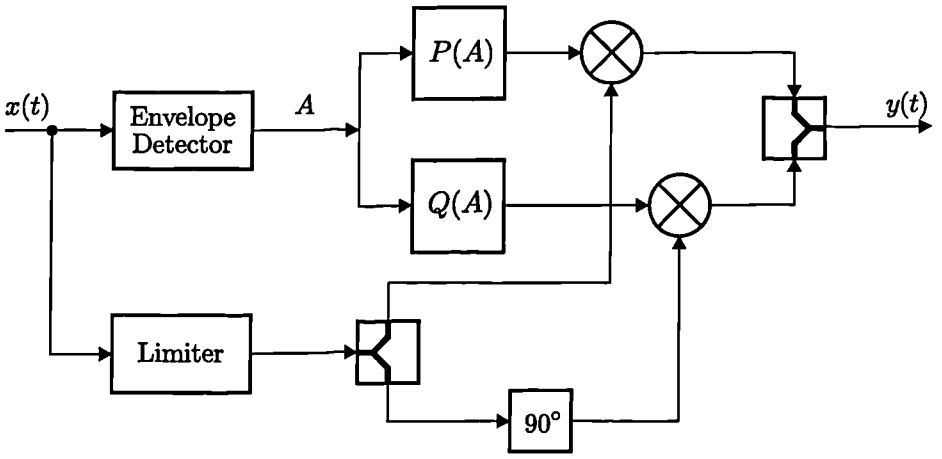
Zero–order

The zero–order TC–PLP is almost identical to its G–PLP counterpart, but has the constraint that the first knot must be zero. If this was not the case, it would not be possible for the predistorter to give zero output amplitude, for zero input amplitude. The transfer characteristic is constant within each segment, *i.e.*, the output is constant even though the input may vary within a given segment. The transfer characteristic is given by

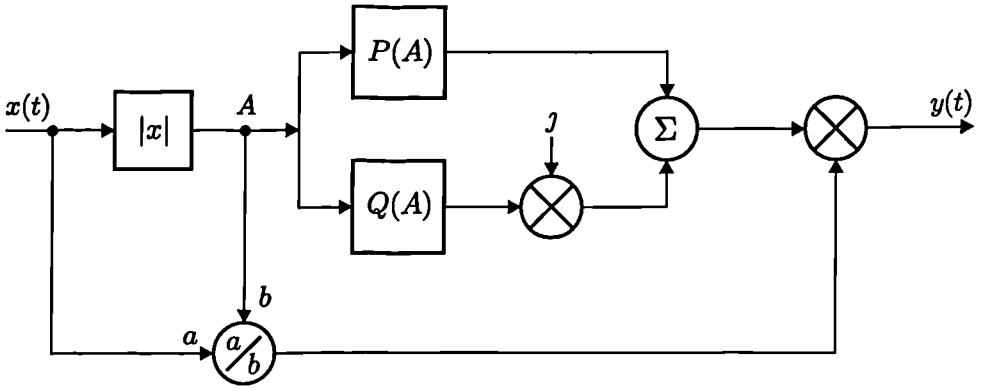
$$T_P(A) = \begin{cases} 0, & 0 \leq A < A[1] \\ T_P(A[1]), & A[1] \leq A < A[2] \\ \vdots & \\ T_P(A[N-1]), & A[N-1] \leq A \leq 1 \end{cases} \tag{6.9}$$

Continuous first–order

The continuous first–order TC–PLP (henceforth simply the continuous TC–PLP) is similar to the continuous G–PLP, but has the added constraint that the transfer characteristic must



(a) Practical configuration



(b) Complex envelope equivalent form

Figure 6.7: Block diagram of the transfer characteristic–type PLP (TC–PLP).

start at zero. This is the only PLP architecture which has been suggested in the literature. The transfer characteristic of the predistorter is

$$T_P(A) = \begin{cases} \left(\frac{A}{A[1]}\right) T_P(A[1]), & 0 \le A < A[1] \\ T_P(A[1]) + \left(\frac{A-A[1]}{A[2]-A[1]}\right)\left(T_P(A[2]) - T_P(A[1])\right), & A[1] \le A < A[2] \\ \vdots & \\ T_P(A[N-1]) + \left(\frac{A-A[N-1]}{1-A[N-1]}\right)\left(T_P(1) - T_P(A[N-1])\right), & A[N-1] \le A \le 1 \end{cases} \tag{6.10}$$

Discontinuous first–order

The discontinuous first–order TC–PLP (henceforth simply the discontinuous TC–PLP) is similar to the discontinuous G–PLP, but is constrained to start at zero. Consequently, the

predistorter transfer characteristic is given by

$$G_P(A) = \begin{cases} \left(\frac{A}{A[1]}\right) T_P(A_U[0]), & 0 \leq A < A[1] \\ T_P(A_L[1]) + \left(\frac{A-A[1]}{A[2]-A[1]}\right)\left(T_P(A_U[1]) - T_P(A_L[1])\right), & A[1] \leq A < A[2] \\ \vdots & \\ T_P(A_L[N-1]) + \left(\frac{A-A[N-1]}{1-A[N-1]}\right)\left(T_P(1) - T_P(A_L[N-1])\right), & A[N-1] \leq A \leq 1 \end{cases} \tag{6.11}$$

6.5 An Initial Adaption Technique for PLP Characteristics

To optimise the linearity of the predistortion amplifier, the knots of the PLP must be carefully chosen. One possible technique is to approximate the transfer characteristic or complex gain of the ideal predistorter, in some way, to generate the 'best' TC–PLP or G–PLP. The optimum characteristic will, however, be dependent upon the input signal p.d.f. and thus it is not obvious what criteria should be used to approximate the characteristic.

In a practical system, the amplifier characteristic will be time–variant, due to environmental factors. The input signal p.d.f. is also often time–variant, *e.g.*, due to the use of differing numbers of channels in a multi–carrier system. Consequently, fixing the PLP characteristic will result in non–optimal performance; however, it may still be acceptable.

The method initially chosen for optimising the PLP characteristic is to use equi–spaced knot abscissæ, and adapt the PLP characteristic during the actual simulation. This requires some time to adapt to the optimal characteristic, but has the advantage of being able to compensate for differing input signals, and time–variant amplifier characteristics. The adaption method is a linear scheme and is similar to that currently being used for adaption of baseband predistorters [14]. It involves the comparison of the ideal predistortion amplifier complex gain or output signal with the actual value, to obtain an error signal which is then used to update the PLP knot values. (In this case, 'optimal linearity' is assumed to occur when the complex gain or transfer characteristics are fully adapted, *i.e.*, the error signal is minimised.)

It will be shown in Section 6.8 that adjusting the knot abscissæ results in improved linearity; however, it is not possible to implement this feature into this simple adaption scheme.

6.5.1 G–PLP Adaption

To adapt the complex gain of the PLP in a particular segment, n, during iteration i, a small multiple of the gain error (the error in gain between the actual predistortion amplifier, and that of the ideal) is added to the previous complex gain for the segment (in this case assumed to be from the previous iteration). Thus, the new complex gain is given by

$$G_{P_L}[i,n] = G_{P_L}[i-1,n] + \xi_L \tag{6.12}$$

$$G_{P_U}[i,n] = G_{P_U}[i-1,n] + \xi_U \tag{6.13}$$

The incremental changes in complex gain, ξ_L and ξ_U, are calculated by multiplying the gain error by some (negative) scaling factors, δ_L and δ_U, thus

$$\xi_L = \delta_L(G_{PA}[i] - G_I) \tag{6.14}$$

$$\xi_U = \delta_U(G_{PA}[i] - G_I) \tag{6.15}$$

where G_I is the ideal predistortion amplifier complex gain, and $G_{PA}[i]$ is the actual predistortion amplifier complex gain during iteration, i, and is simply given by

$$G_{PA}[i] = \frac{y[i]}{x[i]} \tag{6.16}$$

where $x[i]$ is the input signal, and $y[i]$ is the predistortion amplifier output signal, during iteration i. Thus, the new complex gain is given by

$$G_{P_L}[i,n] = G_{P_L}[i-1,n] + \delta_L\left(\frac{y[i]}{x[i]} - G_I\right) \tag{6.17}$$

$$G_{P_U}[i,n] = G_{P_U}[i-1,n] + \delta_R\left(\frac{y[i]}{x[i]} - G_I\right) \tag{6.18}$$

$$\tag{6.19}$$

6.5.2 TC–PLP Adaption

It is possible to use the G–PLP adaption algorithm to optimise the TC–PLP, however, slightly improved linearity can be attained with the use of an alternative algorithm. To adapt the transfer characteristic of the PLP in a particular segment, n, during iteration i, a small multiple of the gain error is added to the previous transfer characteristic for the segment (in this case assumed to be from the previous iteration). Thus, the new transfer characteristic is given by

$$T_{P_L}[i,n] = T_{P_L}[i-1,n] + \xi_L \tag{6.20}$$

$$T_{P_U}[i,n] = T_{P_U}[i-1,n] + \xi_U \tag{6.21}$$

The incremental changes in the transfer characteristic, ξ_L and ξ_U, are calculated by multiplying the gain error by the input magnitude and the (negative) scaling factors δ_L and δ_U, thus

$$\xi_L = \delta_L|x[i]|\left(\frac{y[i]}{x[i]} - G_I\right) \tag{6.22}$$

$$\xi_U = \delta_U|x[i]|\left(\frac{y[i]}{x[i]} - G_I\right) \tag{6.23}$$

Consequently, the new transfer characteristic is given by

$$T_{P_L}[i,n] = T_{P_L}[i-1,n] + \delta_L|x[i]|\left(\frac{y[i]}{x[i]} - G_I\right) \tag{6.24}$$

$$T_{P_U}[i,n] = T_{P_U}[i-1,n] + \delta_U|x[i]|\left(\frac{y[i]}{x[i]} - G_I\right) \tag{6.25}$$

6.5.3 Calculation of scaling factors

The values of the scaling factors, δ_L and δ_U, are dependent upon the PLP configuration. For zero–order and continuous PLP configurations, only the upper breakpoint scaling factor, δ_U, is required (with the lower breakpoint scaling factor, δ_L, equal to zero), since the knot values are constrained. For the discontinuous PLP configurations, both scaling factors are required, as both the upper and lower knot values bounding a segment can be altered.

For the zero–order PLPs, the scaling factor is simply required to be some small–valued negative constant, γ (of the order of -0.01). Consequently,

$$\delta_U = \gamma \tag{6.26}$$

$$\delta_L = 0 \tag{6.27}$$

For the first–order PLPs, adaption is improved by multiplying α by the relative position, in the segment, of the input amplitude. Thus, in a given segment, n,

$$\delta_U = \gamma\left(\frac{A - A[n]}{A[n+1] - A[n]}\right) \tag{6.28}$$

$$\delta_L = 1 - \delta_U \tag{6.29}$$

6.6 Simulation Results

6.6.1 Time Domain Results

The envelope of the predistorter output signal for the various configurations of G–PLPs and TC–PLPs, once fully adapted, are shown in Figure 6.8, using 4 segments. The respective envelopes of the predistortion amplifier output are are shown in Figure 6.9.

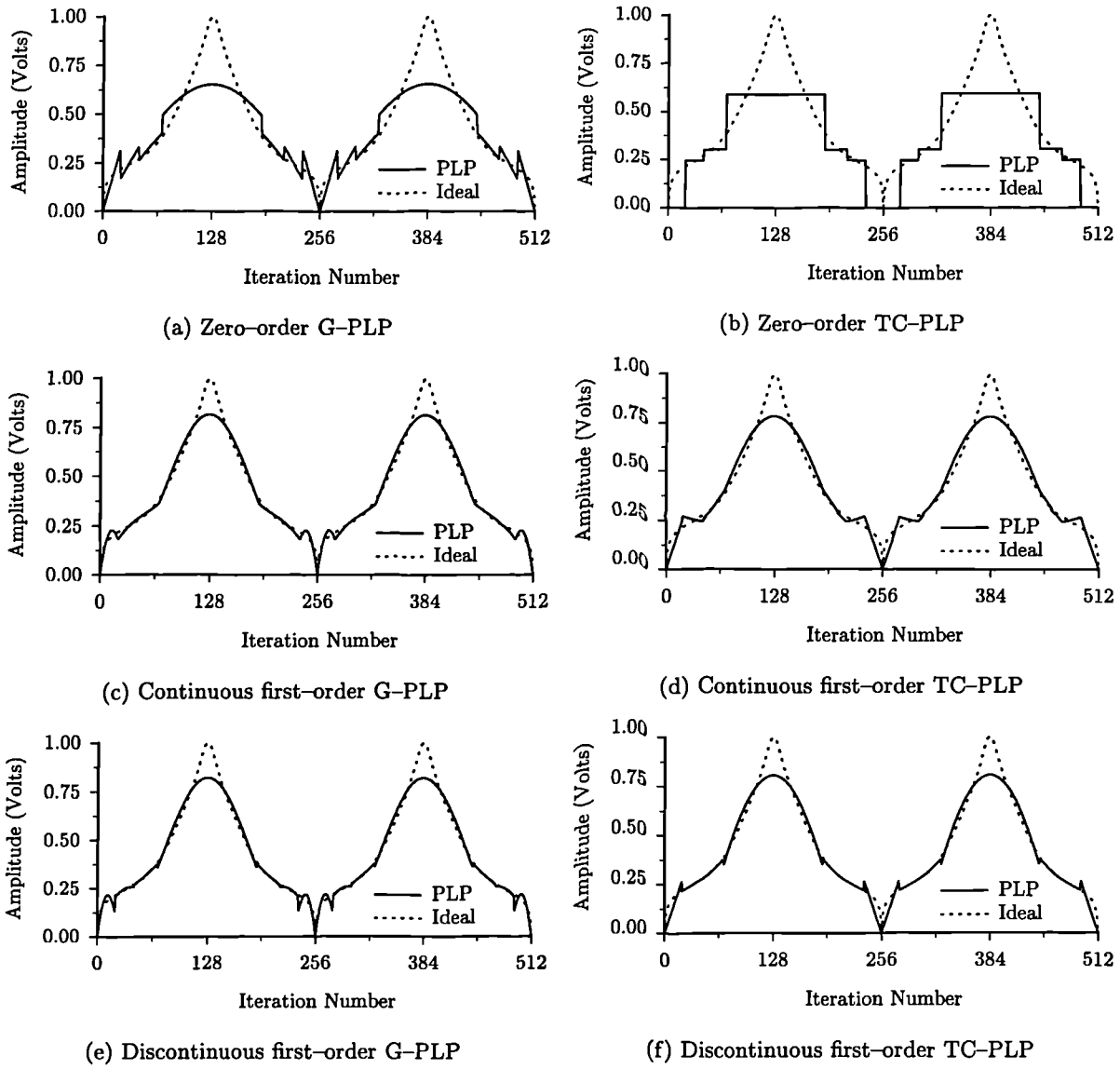
In all cases the peak output of the predistorter is significantly less than that of the ideal predistorter. However, due to the heavily saturated amplifier characteristic, the predistortion amplifier peak output is much closer to the ideal. This decrease in peak output is less significant for the first–order PLPs than for the zero–order, and is also reduced by increasing the number of segments. For first–order PLPs with four or more segments, the reduction in peak output is less than 0.2dB. In the case of zero–order PLPs with four of more segments, the reduction in peak output is below 0.8dB.

The zero–order predistorter envelope characteristics are only crude approximations to the ideal predistorter, and thus the linearity improvement would be expected to be poor. The predistortion amplifier envelope using the zero–order G–PLP exhibits a highly discontinuous nature, with large spikes at low envelope levels. In the TC–PLP case, the predistortion amplifier envelope shows a 'dead–band' at low envelope levels, caused by the added constraint, and the envelope is constant in each segment.

The first–order predistorter envelope characteristics provide a much improved fit to the ideal. The transition between segments is much reduced compared to the zero–order G–PLP, resulting in a reduced spike amplitude at low envelope levels. The difference in envelope between the continuous and discontinuous PLPs is only slight; *i.e.*, the discontinuous PLPs produce envelopes which are almost continuous.

6.6.2 Spectral Response

The predistortion amplifier output spectra, for a two–tone input signal, using 4–segment PLPs, are shown in Figure 6.10. There is a significant improvement in the maximum IMP level compared with the unpredistorted amplifier (Figure 5.7 on page 86). However, the higher–order IMPs are not reduced by the same extent, in fact some have been increased in level. The spectral occupancy of the predistortion amplifier is greater than that of the amplifier alone, due to the increase in the number of frequency components applied to the amplifier input.



(a) Zero–order G–PLP

(b) Zero–order TC–PLP

(c) Continuous first–order G–PLP

(d) Continuous first–order TC–PLP

(e) Discontinuous first–order G–PLP

(f) Discontinuous first–order TC–PLP

Figure 6.8: Envelope of 4 segment PLP output signals.

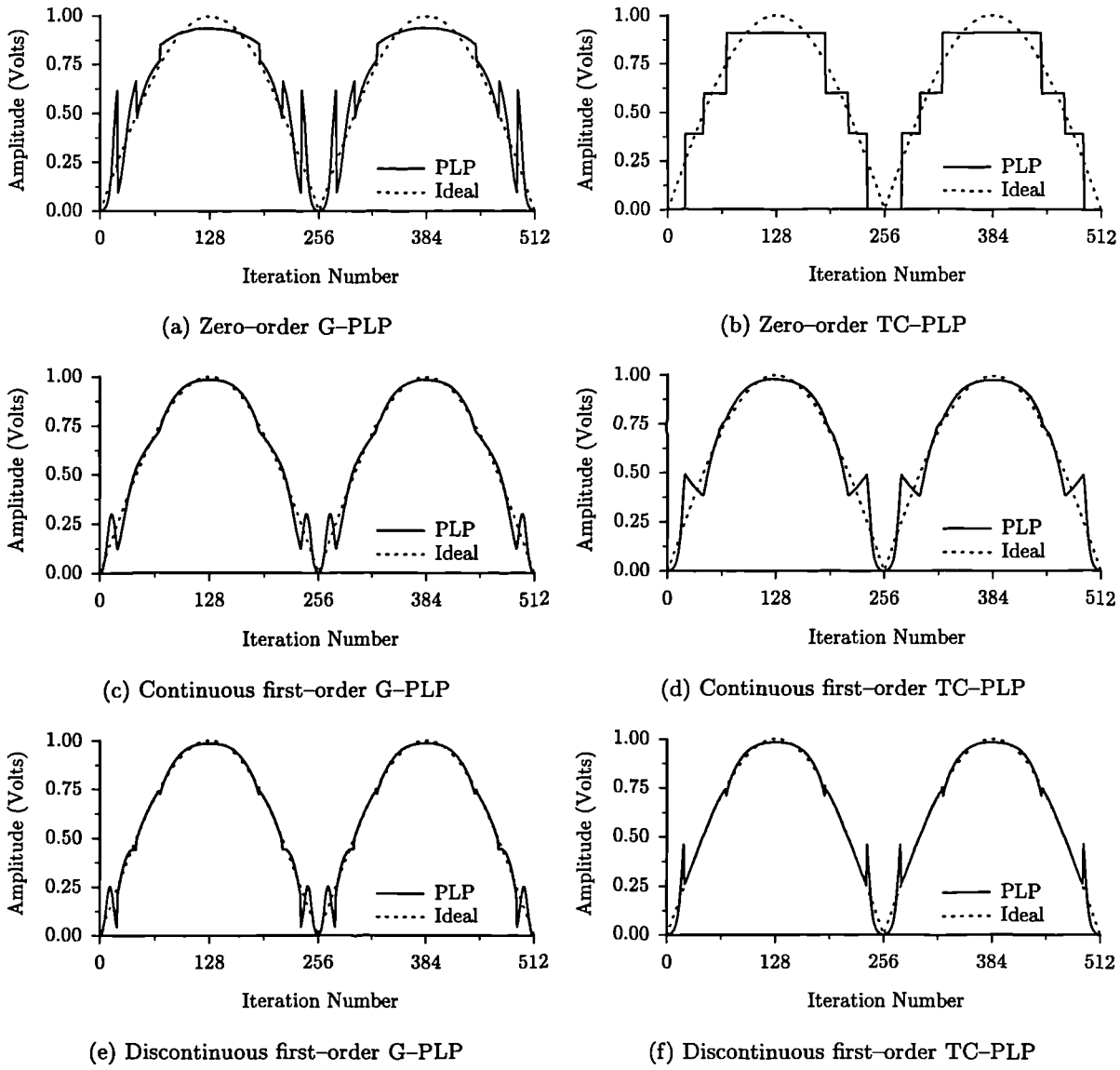
The improvement which can be attained with a 4–segment PLP is significant, and is presented in Table 6.1. It can be seen that greater than 16dB of intermodulation reduction is possible.

PLP Type	G–PLP	TC–PLP
Zero–order	8.7	9.1
Continuous	12.9	9.4
Discontinuous	16.3	16.4

Table 6.1: Improvement in the maximum IMP level (dBc) for 4–segment PLPs.

To highlight the linearity performance of the PLP, the improvement in IMP level utilising differing numbers of segments is shown in Figure 6.11. For all PLP implementations, the linearity improvement generally becomes more significant as the number of segments is increased, since the approximation accuracy of piecewise–linear segments obviously increases with increasing numbers of segments.

It would be expected that the discontinuous PLPs provide the best linearity improvement



(a) Zero–order G–PLP

(b) Zero–order TC–PLP

(c) Continuous first–order G–PLP

(d) Continuous first–order TC–PLP

(e) Discontinuous first–order G–PLP

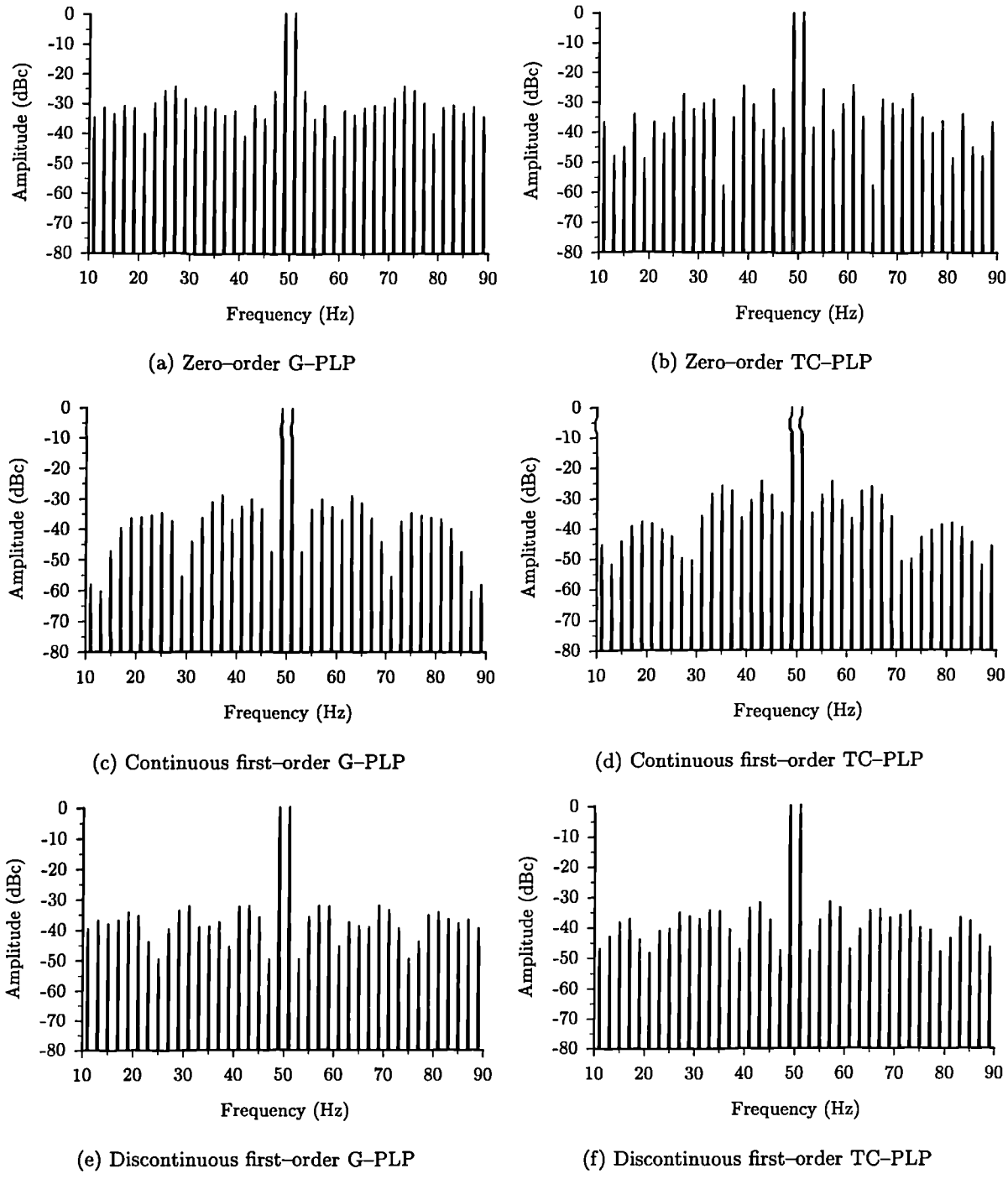
(f) Discontinuous first–order TC–PLP

Figure 6.9: Envelope of predistortion amplifier output signals using a 4 segment PLP.

(since their less constrained nature could provide an improved global approximation), followed by the continuous PLPs, and then the zero–order PLPs. This can be seen to generally be the case, with the exception of the zero–order and continuous TC–PLPs, which produce very similar levels of improvement. It is interesting to note that, although the discontinuous and continuous PLP envelopes are very similar, the linearity performances are significantly different. This suggests that the linearity is highly sensitive to the PLP characteristic.

The discontinuous G–PLP has approximately 3–4dB improved linearity over the continuous G–PLP, which is approximately 3–4dB better than the zero–order G–PLP. The discontinuous TC–PLP exhibits an improvement over the continuous and zero–order TC–PLPs of between 3–9dB.

Comparisons between the G–PLP and the TC–PLP show that the discontinuous G–PLP and TC–PLP have similar levels of performance. The continuous G–PLP is better than the TC–PLP for less than 6 segments, after which it is worse. The zero–order G–PLP is significantly worse than its TC–PLP counterpart.



(a) Zero–order G–PLP

(b) Zero–order TC–PLP

(c) Continuous first–order G–PLP

(d) Continuous first–order TC–PLP

(e) Discontinuous first–order G–PLP

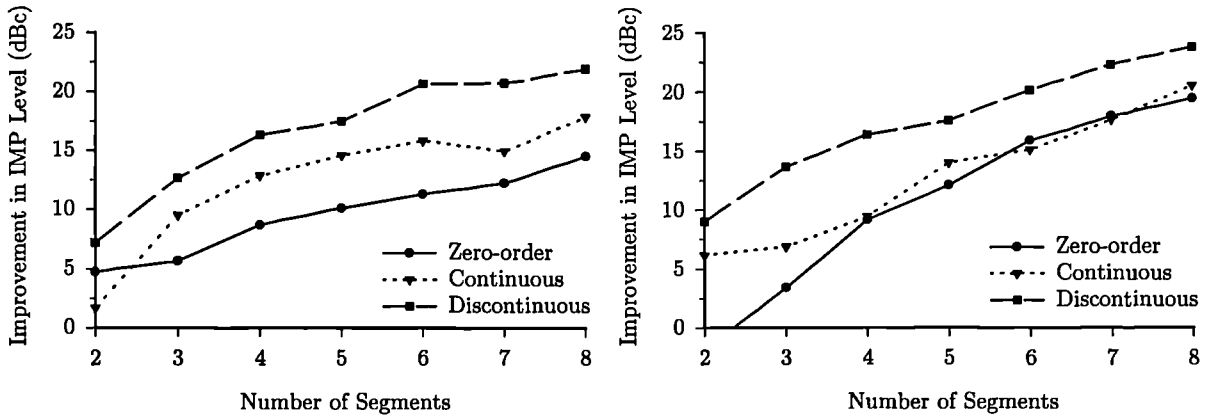
(f) Discontinuous first–order TC–PLP

Figure 6.10: Spectrum of predistortion amplifier output using a 4 segment PLP.

6.6.3 Transfer Characteristics

Although the G–PLP characteristic is chosen with regard to optimising its complex gain, it is instructive to study its transfer characteristic. The same is true for the complex gain characteristic of the TC–PLP, which will be studied in the following section.

The transfer characteristics of the 4–segment G–PLPs are shown in Figure 6.12. The zero–order G–PLP exhibits a transfer characteristic which changes linearly in each segment, and thus has a first–order response. The first–order G–PLPs have a transfer characteristic which



(a) G–PLP (b) TC–PLP

Figure 6.11: Improvement in highest IMP of predistortion amplifier using PLP.

changes in some nonlinear fashion in each segment. It may be simply shown that the characteristic is in fact a quadratic, and thus of second–order.

Whilst this increase in order of the G–PLPs over the TC–PLPs might be expected to give improved approximation to the ideal, this is not necessarily the case. The reason for this is that all of the segments of the transfer characteristics of the G–PLP, if extended, must all pass through zero. This constraint causes their transfer characteristics to exhibit large deviations from ideal; this is particularly noticeable on the zero–order G–PLP response.

The transfer characteristics of the 4–segment TC–PLPs are shown in Figure 6.13. The zero–order TC–PLP shows a crude approximation to the ideal predistorter transfer characteristic, particularly in the first segment; since the output amplitude is constrained to be zero. The first–order TC–PLPs show a much enhanced fit to the ideal predistorter transfer characteristic, due to the use of piecewise–linear, rather than piecewise–constant segments.

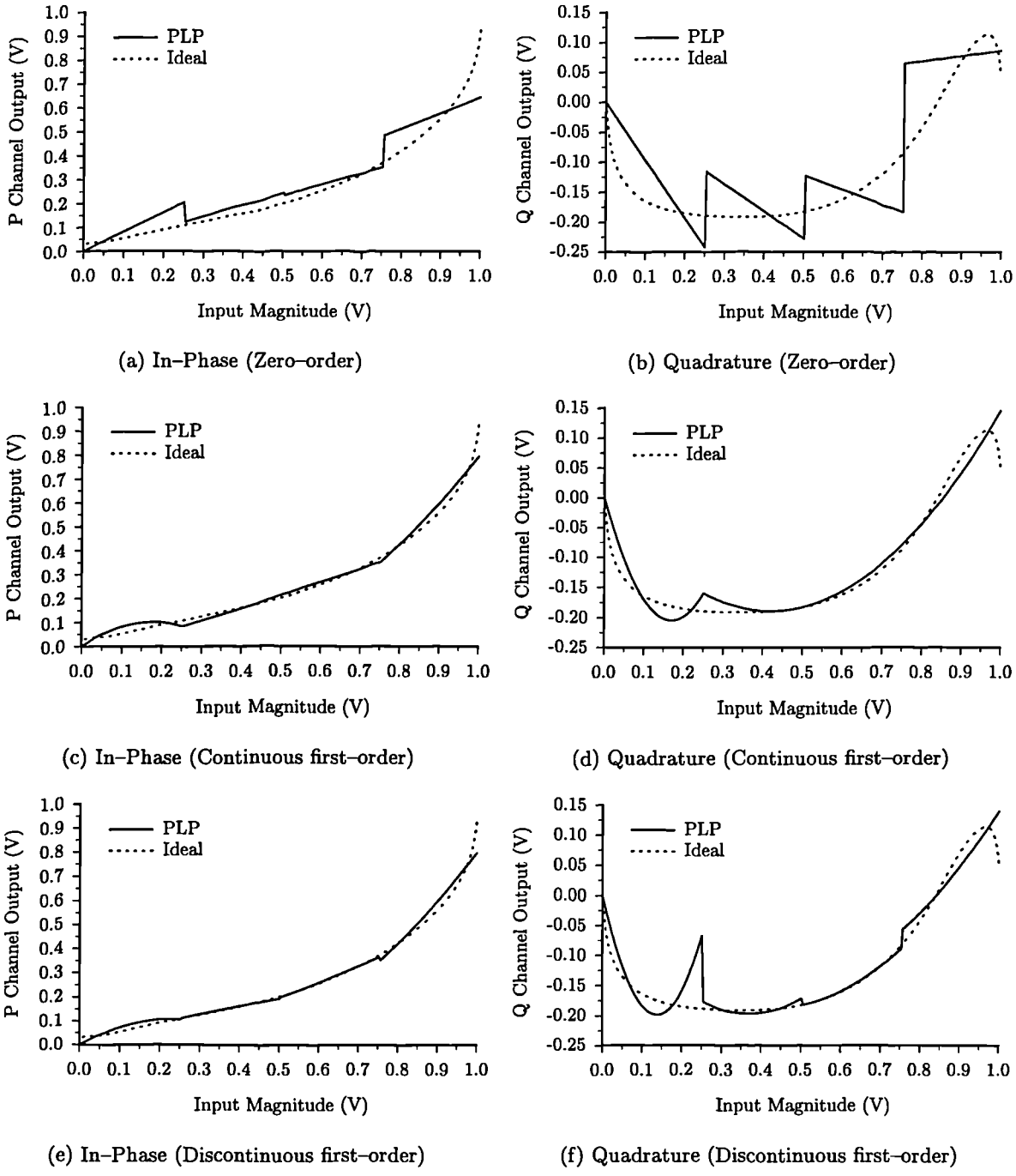
In both the G–PLP and the TC–PLP, the continuous and discontinuous transfer characteristics are very similar.

6.6.4 Gain Characteristic

The complex gain characteristics of the 4–segment G–PLPs are shown in Figure 6.14. The PLP characteristics shows a generally good approximation to the ideal, with the exception of the first segment. The poor fit is due to the large change in gradient (*i.e.*, large second derivative) of the complex gain within this segment.

The complex gain characteristics of the 4–segment TC–PLPs are shown in Figure 6.15. The zero–order TC–PLP generally shows a good fit to the imaginary part of the complex gain, except for first segment, since the transfer characteristic is constrained to be zero in this segment. The approximation to the real part of the complex gain is not as good. The first–order TC–PLPs show an improved fit to the ideal, with the exception of the first segment. It may be simply proved that the complex gain in the first segment is constant. Thus the approximation in the first segment is worse than with the first–order G–PLPs.

In all cases it can be seen that the piecewise fit to the ideal predistorter is not consistently good for all segments. Therefore, altering the knot abscissæ would result in a better global approximation, and hence improved linearity.



(a) In–Phase (Zero–order)

(b) Quadrature (Zero–order)

(c) In–Phase (Continuous first–order)

(d) Quadrature (Continuous first–order)

(e) In–Phase (Discontinuous first–order)

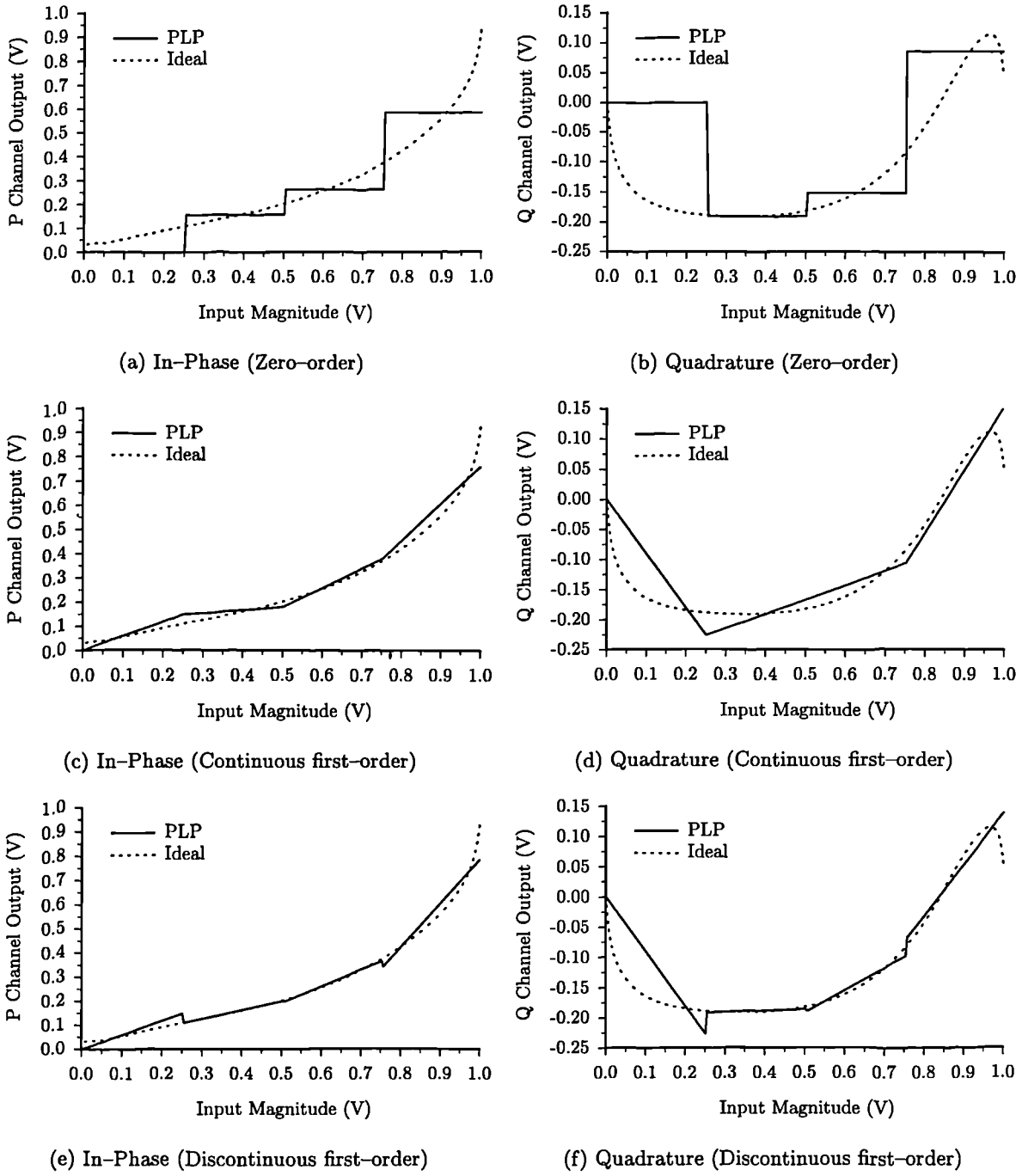
(f) Quadrature (Discontinuous first–order)

Figure 6.12: Transfer characteristic of a 4 segment G–PLP.

6.6.5 Discussion

The simulation results presented illustrate that piecewise–linear predistortion can significantly improve the linearity of an amplifier. However, the improvement in linearity which can be attained may not be related conclusively to other methods in the literature, since no analogue predistortion of class C amplifiers appears to have been published.

It can be seen that the envelope and transfer characteristics of the PLPs show a significant departure from the ideal at low amplitudes. This is due to the rapid change in amplitude and phase of the amplifier transfer characteristic at low amplitude levels. This could be improved



(a) In–Phase (Zero–order)

(b) Quadrature (Zero–order)

(c) In–Phase (Continuous first–order)

(d) Quadrature (Continuous first–order)

(e) In–Phase (Discontinuous first–order)

(f) Quadrature (Discontinuous first–order)

Figure 6.13: Transfer characteristic of a 4 segment TC–PLP.

by having predistorter segments which are not equally spaced. The departure from the ideal envelope at low levels is also apparent, to a lesser extent, in the feedforward amplifier output envelope shown in Figure 4.14(b).

The linearity improvement, in terms of highest IMP, provided by the first–order G–PLPs and TC–PLPs is broadly similar. The zero–order TC–PLP exhibits significantly better performance than the zero–order G–PLP.

For the TC–PLP it has been found that use of the zero–order and continuous configurations cases result in similar levels of linearity improvement. This is a surprising result, given the very crude nature of the zero–order TC–PLP. The fact that the output power does not change

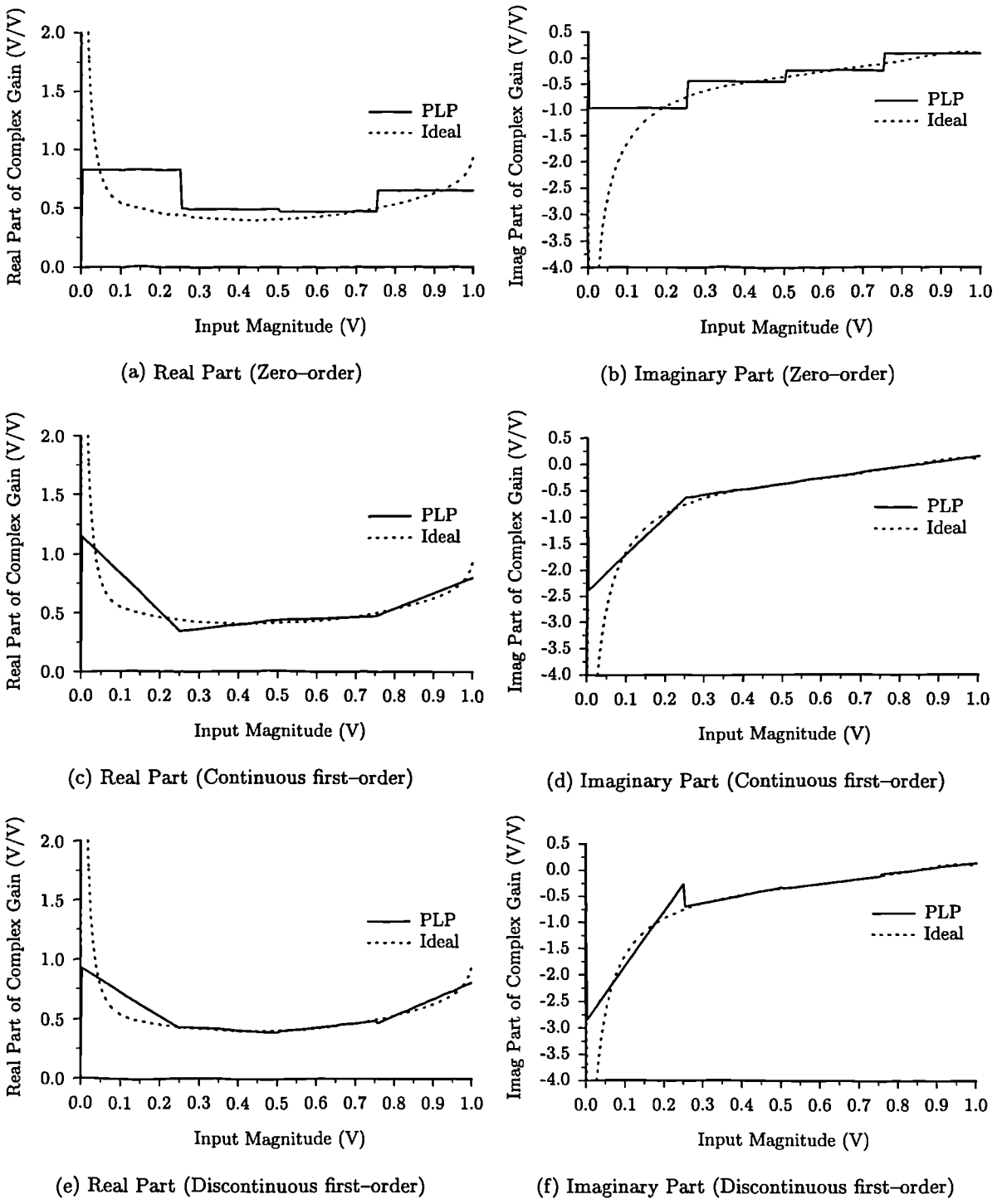
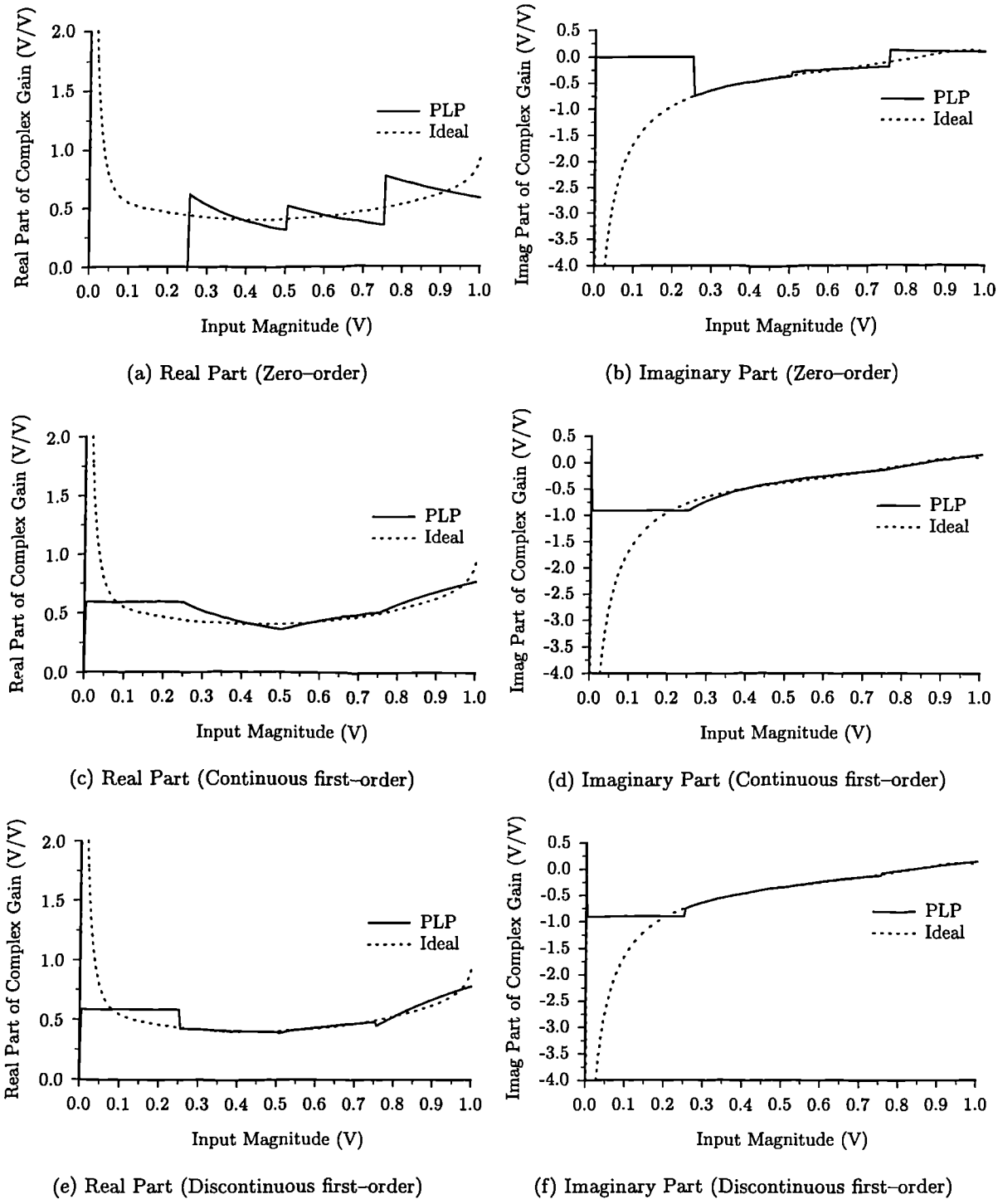


Figure 6.14: Complex gain characteristic of a 4 segment G–PLP.

at all for a change in input power, in each segment, will have linearity implications for an input signal with time–variant characteristics. However in a feedforward system, the error amplifier will supply the extra output power in these regions.

The discontinuous PLPs produce characteristics which are almost continuous. However, their linearity performance is significantly better than for the continuous PLPs. Due to the increase in the number of knots used in a discontinuous PLP (of the order of twice the number required for a continuous PLP), it may be less complex to use a continuous PLP with more segments.



(a) Real Part (Zero–order)

(b) Imaginary Part (Zero–order)

(c) Real Part (Continuous first–order)

(d) Imaginary Part (Continuous first–order)

(e) Real Part (Discontinuous first–order)

(f) Imaginary Part (Discontinuous first–order)

Figure 6.15: Complex gain characteristic of a 4 segment TC–PLP.

6.7 Distortion Measurement Model

In the previous section it has been shown that the use of the improvement in IMP level as the performance metric of a predistorter is not a particularly good one in this application. This is due to the significant level of the high–order IMPs which are generated by the class C amplifier. An improved performance metric may be obtained by measuring the amount of distortion present in the predistortion amplifier output signal.

To obtain a signal consisting almost entirely of distortion, the first loop in a feedforward system can be used (Section 3.7). This method subtracts the amplifier output from a delayed version of the input signal, to obtain an error signal which consists predominantly of distortion. Since the amplifier model is memoryless, it does not contain any frequency–dependence, resulting in zero delay through the simulated amplifier block. Therefore, it is only necessary to equalise the amplitude and phase of the two paths to achieve cancellation.

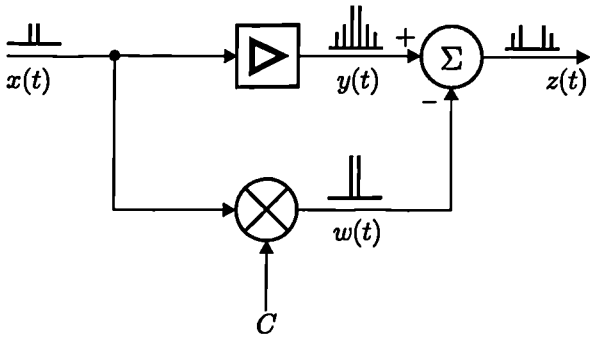


Figure 6.16: Block diagram of distortion measurement method.

The distortion measurement model is shown in Figure 6.16. The output signal will consist of a linear version of the input signal, and distortion components. The distortion components fall into two categories; those at the same frequencies as the input, and those at other frequencies. The ideal cancellation signal would simply cancel the linear part, producing a signal containing all of the distortion. However, such a signal cannot be attained in practice, because the linear signal cannot be isolated from the distortion at the same frequencies. Therefore, the cancellation signal is set at such a level as to generate zero output at those frequencies present in the input spectrum. As a consequence, the distortion signal produced, whilst consisting purely of distortion, does not contain *all* of the distortion. This difference is purely academic in this application; since only a *relative* measure of the amount of distortion is required.

To determine the required gain of the cancellation path, correlation of the distortion signal, $y(t)$, and the input signal, $x(t)$, is performed, in order to produce a dc error signal. Minimisation of this error signal will result in optimal cancellation of those frequency components present in the input signal. However, in complex envelope form, correlation of the input and output signals does not produce a dc component, as will now be shown.

Assume that the input signal contains a component at a frequency, ω_0, with amplitude A. If this component is only partially cancelled, its amplitude at the output will be given by B. Thus, correlation of the two signals in complex envelope form yields

$$(A\cos\omega_0 t + jA\sin\omega_0 t) \times (B\cos\omega_0 t + jB\sin\omega_0 t) = AB(\cos 2\omega_0 t + \sin 2\omega_0 t) \qquad (6.30)$$

which does not have a dc component. To alleviate this problem, it is necessary to correlate the real (or imaginary) part of the input signal with the output signal, thus (for the real part of the input signal)

$$(A\cos\omega_0 t) \times (B\cos\omega_0 t + jB\sin\omega_0 t) = \frac{AB}{2}(1 + \cos 2\omega_0 t + \sin 2\omega_0 t) \qquad (6.31)$$

Hence, the dc component is equal to $\frac{AB}{2}$, and so gives a measure of uncancelled fundamental signal. A digital filter is used to eliminate the harmonic term. The correlation process effectively downconverts the signal to baseband, and so any distortion will also be downconverted. Thus, assuming a two–tone input, the lowest unwanted frequency in the correlation signal

will be at the difference frequency of the two tones. Therefore, it is necessary to filter the correlation signal tightly to eliminate the effect of the distortion. This has been achieved using a second–order elliptic digital filter, designed to reduce unwanted products in the correlation signal by a minimum of 30dB. It is important that the delay through the filter is small, to avoid instability of the adaption algorithm.

A linear adaption algorithm of the same form as that used in the PLP adaption process is used to optimise the cancellation. If the dc component of the correlation signal is given by S_{dc}, the cancellation path gain, C, at iteration i is given by

$$C[i] = C[i-1] + \gamma S_{dc}[i] \tag{6.32}$$

where γ is a small negative constant, whose value depends upon the delay through the digital filter, and is typically of the order of -0.001.

Average and Peak Distortion

Two useful performance metrics can be obtained using the distortion measurement model. The *average distortion* produced by the predistortion amplifier is found by integrating the distortion envelope[2] over one cycle of the input signal envelope. The *peak distortion* is equal to the peak of the distortion envelope within one cycle of the input signal envelope. The improvement in either of these factors is defined as their difference compared to that of the amplifier alone, and is quoted in dB.

Although piecewise–linear predistortion has application in any system requiring analogue predistortion, the particular emphasis in this work is its use as a supplementary technique with feedforward linearisation. This has particular implications for the choice of the performance metric used to optimise the predistorter.

The choice of the metric used is dependent upon the nature of the error amplifier in the feedforward amplifier. The power consumption of a class A amplifier is dependent upon its bias current, which must be larger than the peak output signal amplitude, and is largely unaffected by its input signal. Thus, if a class A error amplifier is used, a reduction in the average distortion of the main amplifier has little effect on the power consumed by the error amplifier, and so the overall efficiency is almost unaffected. However, a reduction in the peak distortion allows the class A amplifier to operate with reduced bias current, and hence overall efficiency is increased. Alternatively, if the bias current is kept constant, the error amplifier will operate more linearly, resulting in improved overall linearity.

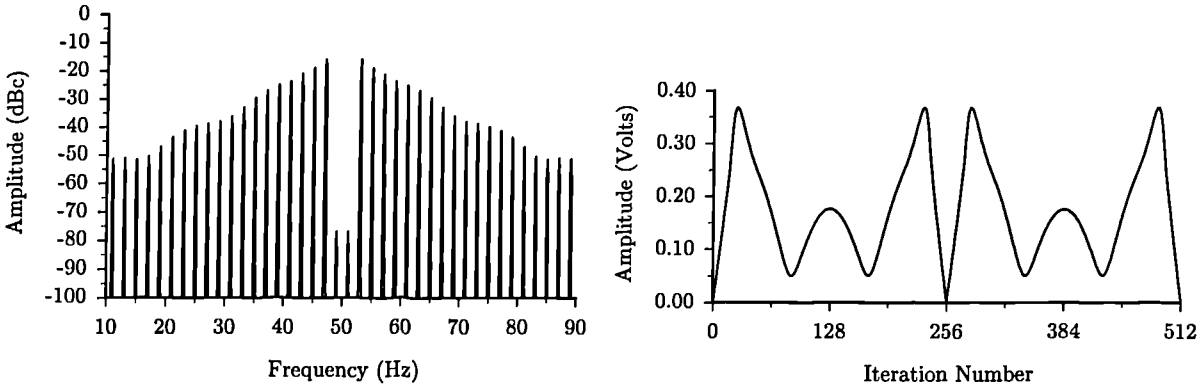
Amplifiers operating in classes AB, B or C, will have a power consumption which is, in part, dependent upon the mean power of the output signal. Therefore, if the error amplifier is not class A, any reduction in the average distortion of the main amplifier will result in increased overall efficiency. A reduction in the peak distortion has no effect on overall efficiency in this case, but will result in improved overall linearity (except if a class C error amplifier is used).

Therefore, in practice it may be necessary to compromise between the two performance metrics in order to achieve optimal performance. It is also important to consider the practical aspects of measuring the peak and average distortion. In principle, both are easy to measure, however, the peak detector required to measure the peak distortion will be susceptible to interference; the average distortion detector will not be as susceptible, due to its averaging nature. Another problem with peak measurements is that, in a multicarrier system, the peak is very dependent upon the relative phases of the channels.

[2]Where the distortion envelope is the envelope of the signal generated by the distortion measurement model, once fully adapted.

6.7.1 Distortion Measurement Model Results

The spectrum of the distortion signal produced using the amplifier is shown in Figure 6.17(a). It can be seen that the input signal component has been cancelled by approximately 75dB, and hence the signal consists almost exclusively of the distortion generated by the amplifier.



(a) Spectrum (b) Envelope

Figure 6.17: Spectrum and envelope of distortion produced by the amplifier.

The envelope of the amplifier distortion signal is shown in Figure 6.17(b), and has a peak value of 0.368V. The distortion signal envelope reduces to near zero, and then increases. It may not be stated that the minimum point on the distortion envelope corresponds to a point on the amplifier transfer characteristic which generates minimum distortion, for two reasons. Firstly, the distortion envelope does not contain all of the distortion. Secondly, the distortion produced by the amplifier is due to the variation in the input envelope, and so no point on it may be isolated in this fashion.

The distortion envelopes of the predistortion amplifier using the various 4–segment PLP configurations are shown in Figure 6.18. It can be seen that, with the exception of the zero–order G–PLP, the peak distortion envelope has been significantly reduced. The distortion envelopes of the predistortion amplifier exhibit more peaks than the amplifier alone.

The improvements in the average and peak distortion levels using differing numbers of segments are shown in Figures 6.19 and 6.20 respectively. It can be seen that there is a general improvement in performance as the number of segments is increased.

The improvement in average distortion follows a similar pattern to that of the improvement of the highest IMP level (Figure 6.11), with little difference in performance between the G–PLP and the TC–PLP. However, the improvement in peak distortion follows a very different pattern. In the G–PLP case, the continuous PLP exhibits the best performance, rather than the discontinuous PLP. The zero–order G–PLP produces only a very limited improvement in the peak distortion. In the TC–PLP case, the discontinuous PLP still offers the best performance, but to a lesser degree than was the case for the other performance metrics. The zero–order TC–PLP exhibits a much better reduction in the peak distortion than its G–PLP counterpart. However, it must be noted that the PLP characteristic was not optimised to reduce either of these metrics to a minimum level, and so any comparisons made cannot be a conclusive guide to the potential linearity performance.

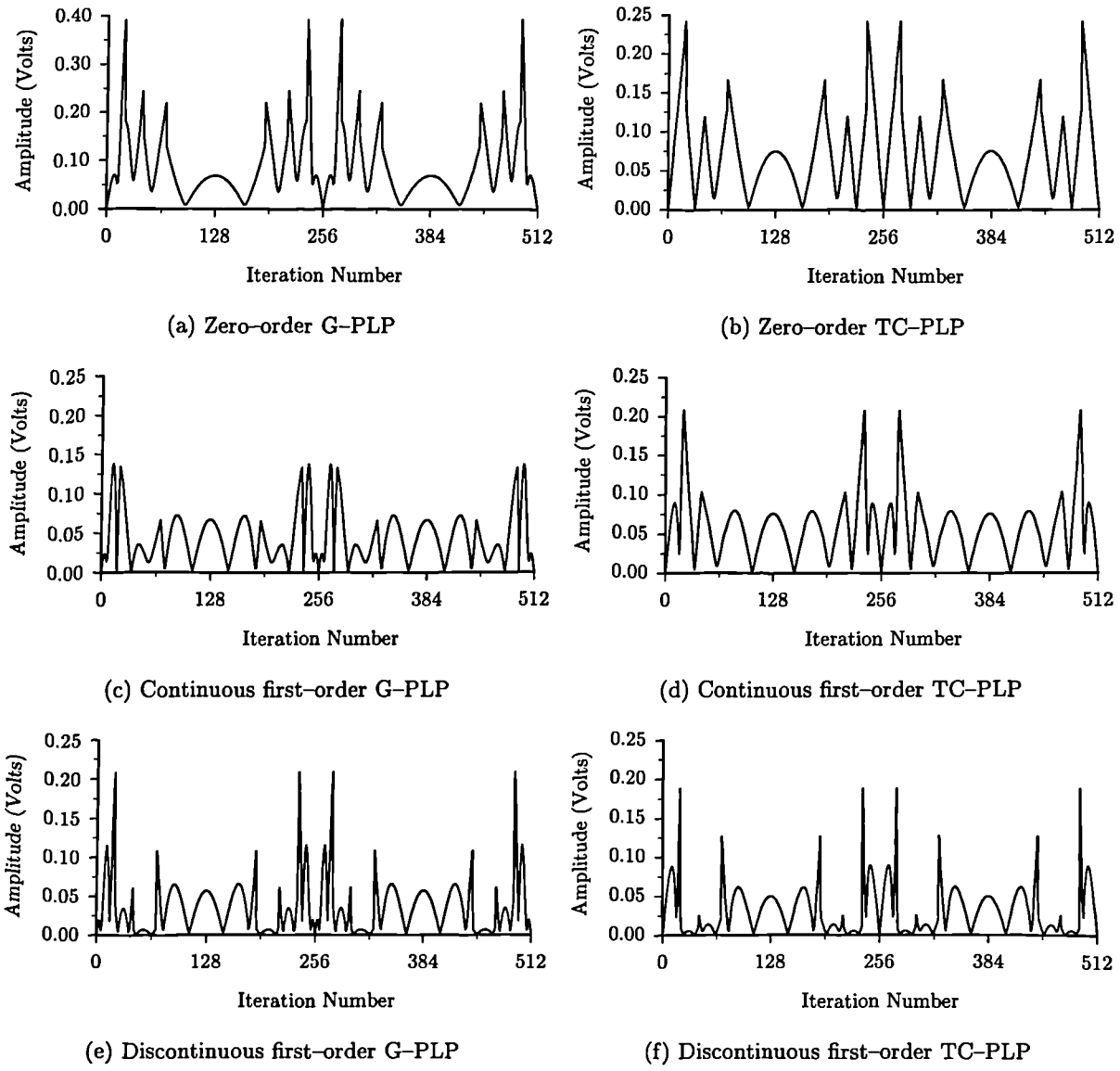
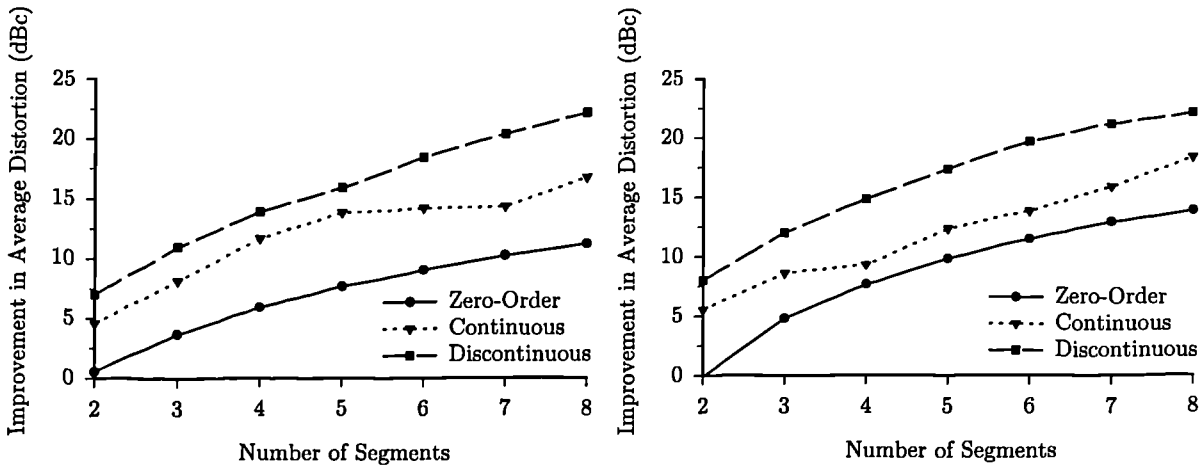


Figure 6.18: Envelope of distortion of predistortion amplifier output using a 4 segment PLP.

6.8 Optimal Least–Squared Fit of Ideal Predistorter Characteristic

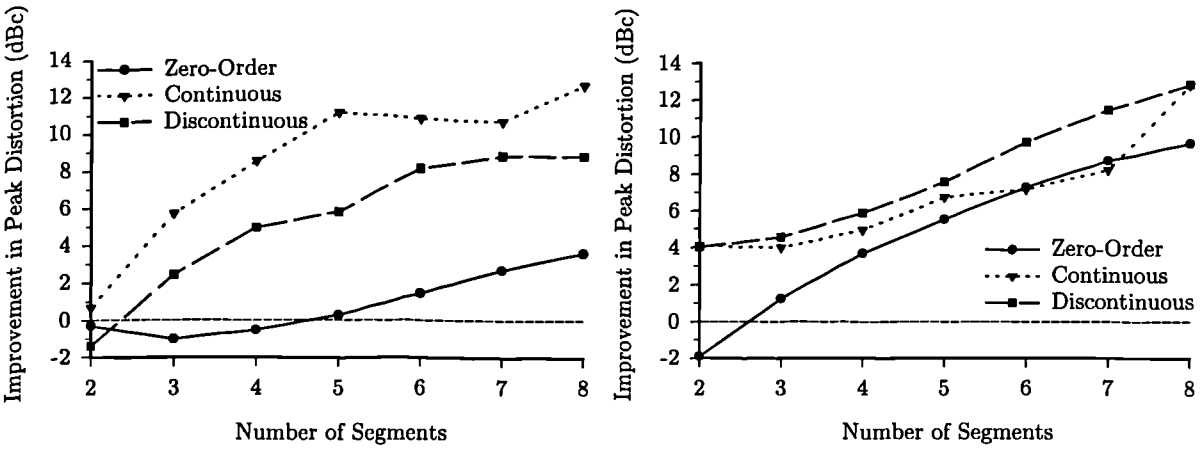
The adaption algorithm used in the previous section assumed that the knots were uniformly spaced. However, this is unlikely to be the optimal separation in general. This is highlighted by the fact that some segments of the PLP produce a better fit than others to the predistorter characteristic. A mechanism is required which allows non–uniform spacing of the knots, to achieve improved linearity. Using the previous adaption algorithm it is not possible to implement optimisation of the knot spacings in an efficient manner. Therefore, an alternative mechanism is proposed.

Since the ideal predistorter characteristic is known, it is possible to fit a piecewise–linear function to it, externally to the simulation process. This could be done using least–squared or minimax fitting procedures. Whilst this would not provide a method of optimising the predistorter in a practical system, since the amplifier characteristic is assumed time–invariant,



(a) G–PLP (b) TC–PLP

Figure 6.19: Improvement in average distortion of predistortion amplifier using PLP.



(a) G–PLP (b) TC–PLP

Figure 6.20: Improvement in peak distortion of predistortion amplifier using PLP.

it would give an improved knot spacing for the predistorter, which can then be used in another optimisation method (such as the one previously described). Since no account is taken of the p.d.f. of the input signal, a suboptimal solution results.

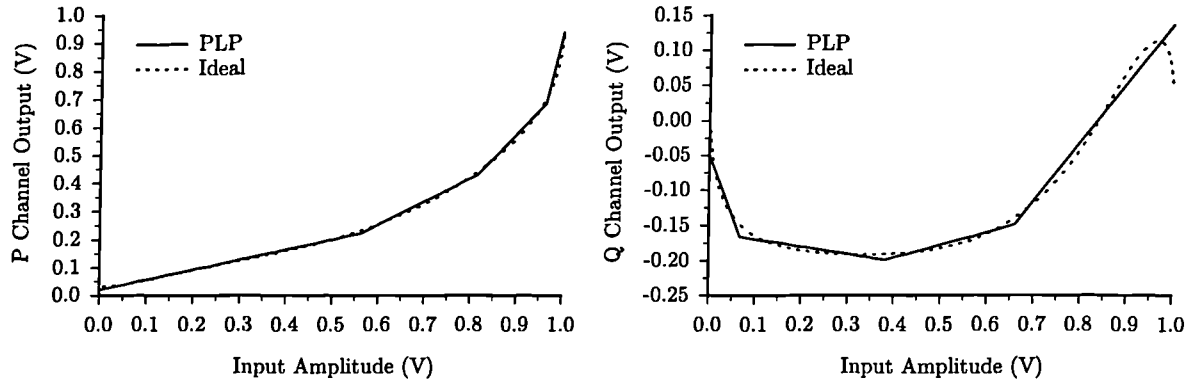
The fitting procedure is complicated by the points of inflexion on the ideal predistorter characteristics (*i.e.*, those points with zero second derivative). No general algorithm appears to have been derived for a minimax approximation with free knots, however, recently a number of algorithms have been published for the case of a least–squared approximation with free knots [15–17]. Baines [16] has designed an algorithm which finds the best discontinuous piecewise–constant or piecewise–linear fit to a continuous function. However, it is noted by Chui *et al* [17] that in the least–squared sense, a one–dimensional function is best approximated by a *continuous* piecewise fit, and that a discontinuous fit converges to a continuous one (continuous piecewise functions being a subset of discontinuous ones). This is validated in this application since it has been shown in the previous section that the discontinuous PLP characteristics were very similar to the continuous ones.

The algorithm used in the fitting procedure is that proposed by Loach and Wathen [15][3], and performs a least–squared fit of a continuous piecewise–linear function. The requirement for the TC–PLP to start at zero has not been taken into account. Whilst it may be possible to extend this algorithm to allow fitting with a piecewise–constant function, so that approximation of a zero–order PLP can be achieved, this has not been attempted in this work.

6.8.1 Results

The transfer characteristic of the optimal least–squared fit 4 segment TC–PLP is depicted in Figure 6.21. The piecewise–linear approximation is significantly better than that attained using equally spaced knots (Figures 6.13(c) and 6.13(d) on page 116). Whilst the PLP does not start at zero, it is close to zero in both the in–phase and the quadrature channels, and so forcing this constraint should not significantly affect the linearity. The linearity figures to be presented for the TC–PLP are derived using the optimal least–squared characteristic with this constraint enforced.

The complex gain characteristic of the optimal least–squared fit 4 segment G–PLP is depicted in Figure 6.22. It may be seen that, as with the TC–PLP, the piecewise–linear approximation is significantly better than that attained using equally spaced knots (Figures 6.14(c) and 6.14(d) on page 117).



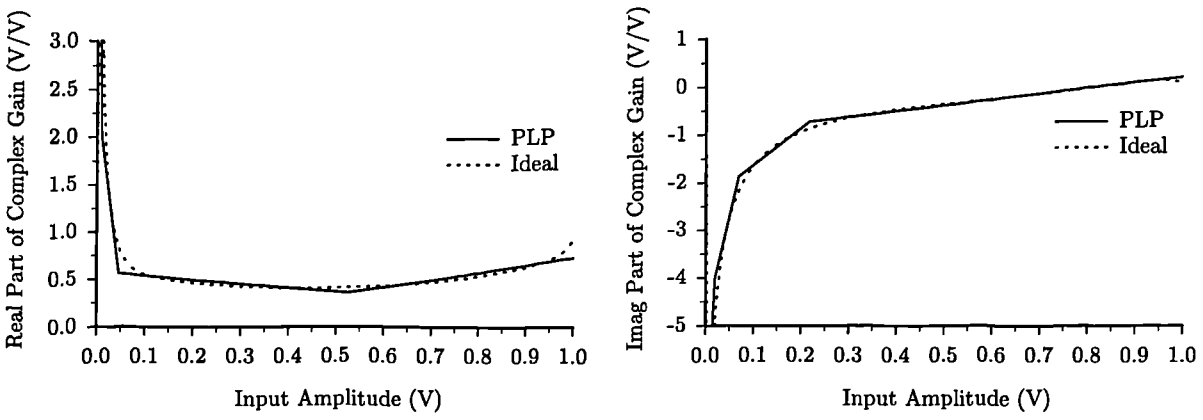
(a) In–phase (b) Quadrature

Figure 6.21: Transfer characteristic of an optimal least–squared 4 segment TC–PLP.

Figure 6.23 shows the residual least–squared error between the ideal predistorter characteristic and the piecewise–linear approximation. It can be seen that the error is reduced by increasing the number of segments, but to a lesser extent each time. The error for the G–PLP is larger than that for the TC–PLP; this is due to the larger range of complex gain (−8.5 to 0.1 for the imaginary part) compared with that of the transfer characteristic (0 to 1 for the P channel).

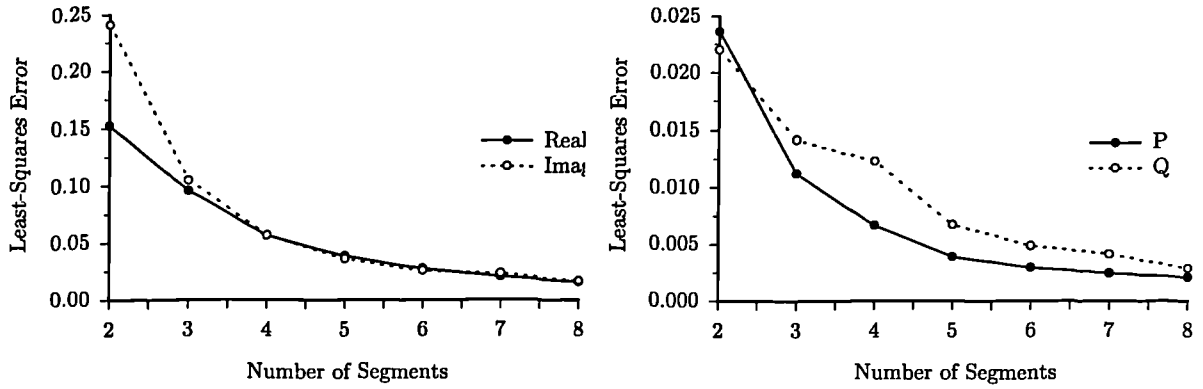
The least–squared error for the TC–PLP is higher for the Q channel, than for the P channel, even though it has a smaller range. This is due to the difference in form of the two channel characteristics. The Q channel characteristic is more complex than the P channel, with two points of inflexion, and so is less well approximated. This is not the case with the G–PLP characteristic, since the forms of the real and imaginary parts of the complex gain function are very similar, resulting in very similar least–squared errors.

[3]The author gratefully acknowledges the assistance of Dr. A. Wathen of the School of Mathematics, University of Bristol, for his provision of this algorithm.



(a) Real Part (b) Imaginary Part

Figure 6.22: Complex Gain of an optimal least–squared 4 segment G–PLP.



(a) G–PLP (b) TC–PLP

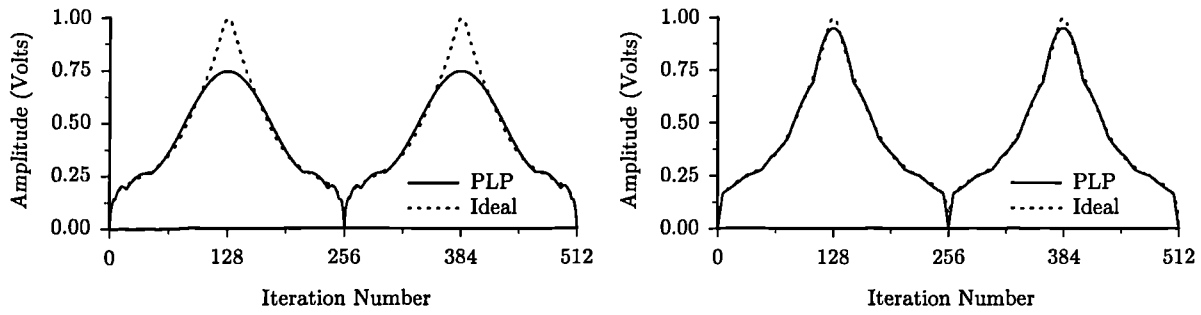
Figure 6.23: Optimal least–squared error for PLPs.

The envelope of the 4–segment PLP outputs, and predistortion amplifier outputs using the optimum least–squared fit characteristics are shown in Figures 6.24 and 6.25. These envelope characteristics have been generated without using any further adaption of the PLP.

It can be seen that the envelope generated by the TC–PLP is much closer to the ideal, compared to its equi–spaced equivalent (Figure 6.8(d) on page 111). Thus, the predistortion amplifier output using the TC–PLP is very similar to the ideal. The spikes in the output, evident in the equi–spaced envelope, are no longer present, and the peak output is closer to that required.

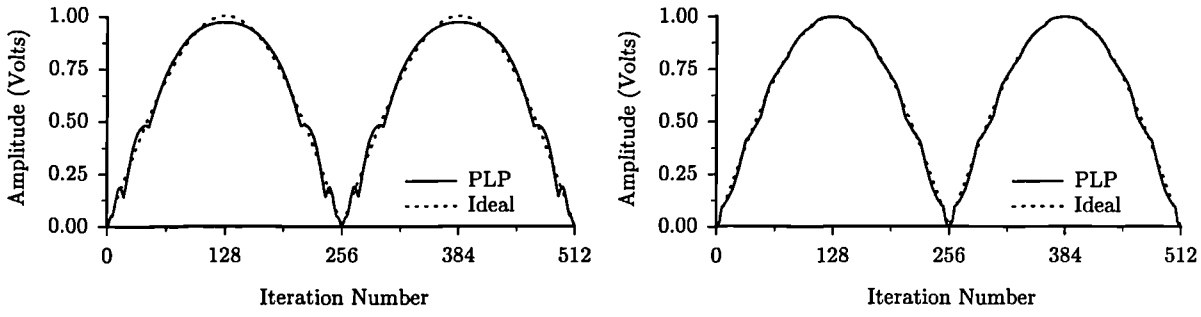
In contrast, the envelope generated by the G–PLP does not appear to be a significant improvement compared with the equi–spaced G–PLP (Figure 6.8(c) on page 111). The predistortion amplifier output using the G–PLP exhibits less significant spikes than present in the equi–spaced G–PLP (Figure 6.9(c) on page 112), although the peak output is still slightly lower than ideal.

The linearity improvements which can be attained using the optimal least–squared fit PLPs are depicted in Figure 6.26. To allow comparison with the equally spaced PLPs, the discontinuous PLP linearity is also shown.



(a) G–PLP (b) TC–PLP

Figure 6.24: Envelope of 4 segment continuous PLP output signals using optimum least–squared characteristics.



(a) G–PLP (b) TC–PLP

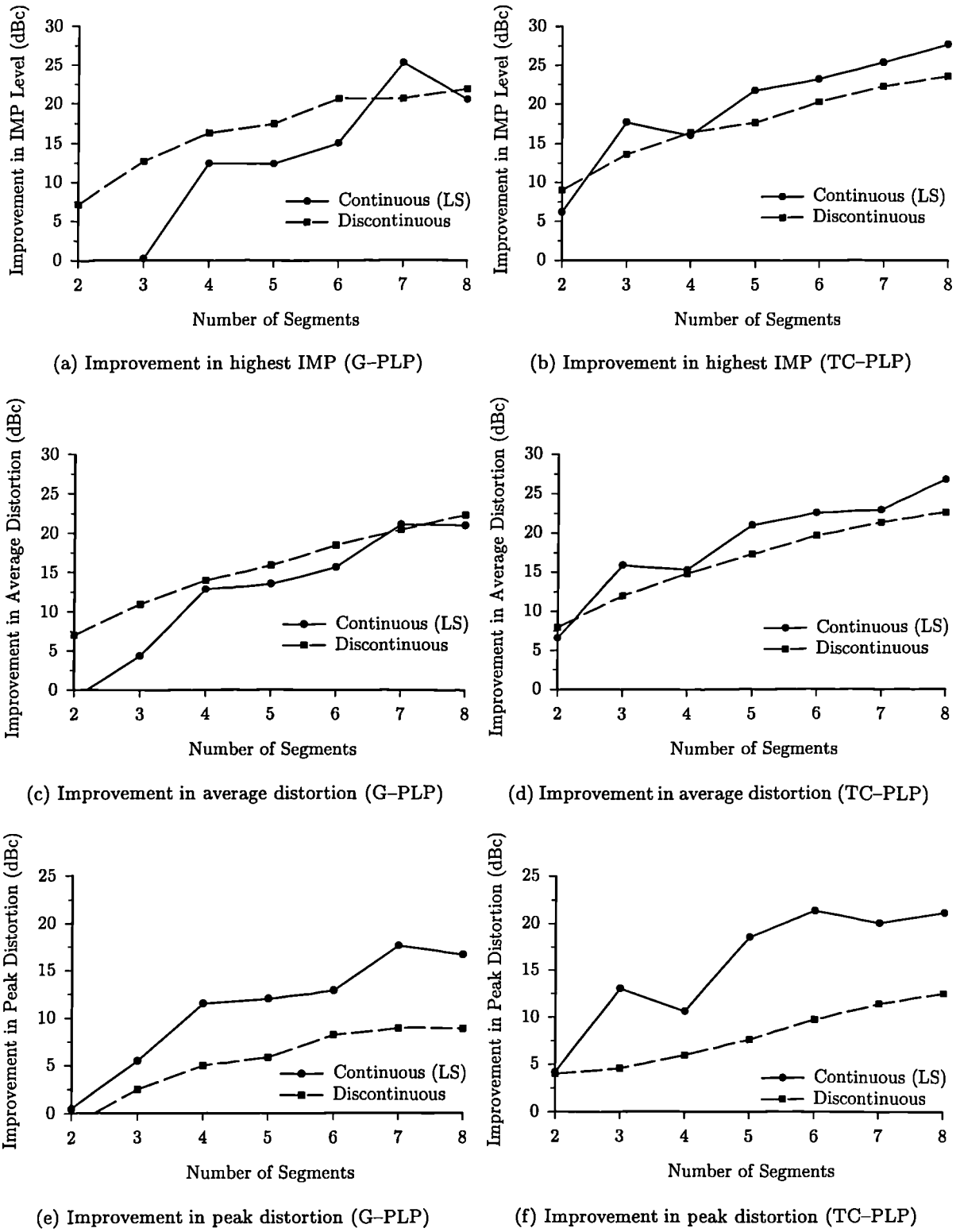
Figure 6.25: Envelope of predistortion amplifier output signals using 4 segment continuous PLPs with optimum least–squared characteristics.

It can be seen that the TC–PLP generally exhibits improved linearity compared with the equi–spaced TC–PLP; this is particularly evident for the peak distortion measurement, which is up to 10dB better. However, although the least–squared error is continually reduced as the number of segments is increased, this does not necessarily lead to improved linearity. This is emphasised by the fact that the linearity for the 4 segment TC–PLP is worse than that for the 3 segment PLP. Therefore, although the optimal least–squared fit is the best global fit to the actual characteristic, it is not necessarily the best fit in terms of linearity for the given input signal.

The optimal least–squared fit G–PLP does not generally exhibit improved linearity (in terms of highest IMP and average distortion) compared with the discontinuous equi–spaced G–PLP; although it is better than the continuous equi–spaced G–PLP. However, in terms of peak distortion, the performance is a significant improvement. This highlights the fact that, although the characteristic appears to be a good approximation, it is necessary to further optimise the performance, taking into account the relevant metric. This optimisation procedure is considered in the following section.

6.9 An Improved Optimisation Scheme

The adaption of a PLP can be considered as an N–dimensional optimisation of a function, where N depends upon the predistorter type, and the number of knots. The goal is to find a value of the variable vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ which minimises the objective function,



(a) Improvement in highest IMP (G–PLP)

(b) Improvement in highest IMP (TC–PLP)

(c) Improvement in average distortion (G–PLP)

(d) Improvement in average distortion (TC–PLP)

(e) Improvement in peak distortion (G–PLP)

(f) Improvement in peak distortion (TC–PLP)

Figure 6.26: Linearity improvement of predistortion amplifier using PLP with optimum least–squared characteristic.

$f(\mathbf{x})$. The variable vector consists of all of the predistorter knots which are required to be optimised. The optimisation would ideally find the global minimum of the distortion metric, and thus the best possible solution. However, this is generally much more complicated than finding a local minimum.

6.9.1 Local Optimisation

There are a large number of algorithms which can find a *local* minimum of an N–dimensional function. These local optimisation schemes will also find the global minimum if a good initial vector, $\mathbf{x}^0$, is used, or if only one minimum is present (in which case the function is said to be *convex*). Given the large number of dimensions possible in this application, it is likely that many local minima will exist. The initial vector chosen is that of the optimal least–squared fit to the nonlinear functions, presented in the previous section.

Many local minimisation algorithms require knowledge of the derivative of the function, which is impractical in this application. This limits the number of algorithms which may be used to *direct–search* algorithms. The two most popular algorithms of this type are the *Downhill Simplex Method* and *Powell's Conjugate Direction Method.*

The Downhill Simplex Method [18–20] (not to be confused with the Simplex Method of Linear Programming) uses $N+1$ points in N dimensions, which form the vertices of a *simplex.* The function is calculated at each point in the simplex, and the vertex which generates the highest function value is replaced by another, using projection of the other vertices. This new simplex is tested, and the procedure repeated, until convergence is reached, and the local minimum found. The method is generally quite slow, but very robust.

Powell's Conjugate Direction Method [21,22] starts at a point in N–dimensional space, and minimises the function along a given direction. The direction is then changed, and minimisation repeated, until convergence occurs. The directions chosen determine how rapidly convergence is attained, and there are a number of criteria which are used in the algorithm to produce the 'best' direction from any point in which to minimise the function. This method is generally the quickest direct search algorithm, but is not as robust as the Downhill Simplex Method.

Both of these techniques have been used in the simulation to enable comparison of their relative performance. However, whilst the Powell's Conjugate Direction Method proved to be generally quicker than the Downhill Simplex Method, it would not always converge to a solution. Therefore, the Downhill Simplex Method has been used exclusively to generate the results.

6.9.2 Optimisation Constraints

The local optimisation schemes described in the previous section are, in common with most techniques, unconstrained in nature, allowing the variable vector $\mathbf{x}$ to take any value. This is unacceptable in this application for a number of reasons, and so constraints must be added.

Knot Entanglement

The abscissæ of the knots are required to satisfy

$$A[0] = 0 < A[1] < A[2] < \cdots < A[N-2] < A[N-1] < A[N] = 1 \tag{6.33}$$

In an unconstrained optimisation scheme it is highly likely that eventually knot entanglement will occur, in which, for a given segment, n,

$$A[n-1] > A[n] \tag{6.34}$$

Whilst it is possible to simply reorder the knots, it is unclear what should be done if a knot abscissa is given a value outside of the two extremes. An improved procedure has been derived

in which the knot abscissæ are replaced in the variable vector by an equivalent form, for which knot entanglement may be more easily avoided.

The abscissa of a given knot n is given by $A[n]$, and is bounded by knots on both sides with abscissæ $A[n-1]$ and $A[n+1]$ (except for the two knots at the extremes). The fractional distance of this knot abscissa to those on either side, $\sigma[n]$, is given by

$$\sigma[n] = \frac{A[n] - A[n-1]}{A[n+1] - A[n-1]} \tag{6.35}$$

Therefore,

$$\begin{aligned} A[0] &= 0 \\ A[1] &= \sigma[1](A[2] - A[0]) + A[0] \\ &\vdots \\ A[N-1] &= \sigma[N-1](A[N] - A[N-2]) + A[N-2] \\ A[N] &= 1 \end{aligned} \tag{6.36}$$

The variable vector is initialised with the relevant σ values, found from the original knot abscissæ. During the optimisation procedure knot entanglement will occur if any σ is outside of the range $0 < \sigma < 1$. If this occurs, the σ value is forced back inside this range by firstly finding its absolute value modulo two (thus forcing the value to be between 0 and 2), and if greater than 1, subtracting from 2.

The knot values cannot be directly found from the σ values, due to its ratiometric nature. However, by introduction of another variable, $\varepsilon[n]$, where

$$\varepsilon[n] = \frac{\sigma[n]}{1 - \varepsilon[n-1](1 - \sigma[n])} \tag{6.37}$$

The knot abscissæ become

$$\begin{aligned} A[0] &= 0 \\ A[1] &= \varepsilon[1]A[2] \\ &\vdots \\ A[N-1] &= \varepsilon[N-1]A[N] \\ A[N] &= 1 \end{aligned} \tag{6.38}$$

It is therefore possible to iterate forward through the list of σ to find the ε values using Equation 6.37 (assuming $\varepsilon[0] = 0$), and then iterate backwards using Equation 6.38 to find the knot values.

Peak Predistortion Amplifier Power

The optimisation procedure is required to minimise the average or peak distortion present in the predistortion amplifier output. If an unconstrained optimisation procedure is used, it has been found that the knot values are continually reduced such that the amplifier is backed–off, and thus the distortion reduced. This occurs even with the class C amplifier model, since although the distortion increases as the amplifier is backed–off in *relative* terms (*i.e.*, compared with the linear output), it may still reduce in *absolute* terms. In order to stop this effect it is necessary that the peak output power of the predistortion amplifier remain

constant. This is also desirable in order to keep the gain through the predistortion amplifier constant.

The PLP characteristics use the quadrature representation of a memoryless nonlinearity, and hence the coefficients are stored in cartesian format. To implement this constraint the highest knot value is converted to polar form. The amplitude is kept at unity, whilst the phase is included in the variable vector $\mathbf{x}$ and applied to the optimisation scheme. After each optimisation stage the amplitude and phase are then converted back into cartesian form to produce the required real and imaginary parts of the PLP characteristic.

Average Predistortion Amplifier Output Power

Whilst the above constraining mechanism is able to maintain the peak output power at all times, it has been found that this is not enough to stop the optimisation scheme causing the amplifier to back–off. The abscissa of the second–highest knot tends to converge to that of the highest knot (unity), allowing the amplifier to be backed–off over most of its range, but producing a sharp distortion peak at peak input amplitude. This effect is highlighted in Figure 6.27, which shows the envelope of the predistortion amplifier without this constraint being used. This problem occurs particularly for the average distortion optimisation, since this distortion peak makes a negligible difference to the average distortion.

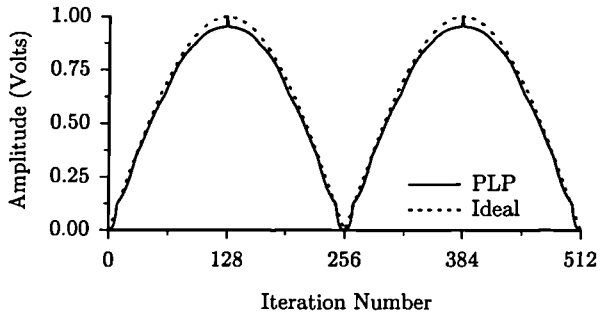


Figure 6.27: Envelope of predistortion amplifier output using 4 segment continuous TC–PLP, after optimisation without constraining the average output power.

To reduce this effect, the *average* output power of the predistortion amplifier should be kept (at least approximately) constant. The method chosen to implement this constraint is to use a *penalty function* [19]. This technique alters the objective function $f(\mathbf{x})$ such that a penalty is added if the constraint is not met, and applies this new function to the unconstrained optimisation routine. The transformed function, $p(\mathbf{x}, R)$, is given by

$$p(\mathbf{x}, R) = f(\mathbf{x}) + \Omega(R, h(\mathbf{x})) \tag{6.39}$$

where the constraint (an equality constraint in this instance) is $h(\mathbf{x})$, and R is is set of penalty parameters. The penalty term chosen is given by

$$\Omega(R, h(\mathbf{x})) = \frac{h(\mathbf{x})^2}{R} \tag{6.40}$$

and is known as a parabolic penalty. The constraint term, $h(\mathbf{x})$, is given by

$$h(\mathbf{x}) = P_{o,ID} - P_{o,ACT}(\mathbf{x}) \tag{6.41}$$

where $P_{o,ID}$ is the ideal average output power, and $P_{o,ACT}$ is the actual average output power of the predistortion amplifier.

The transformed function, $p(\mathbf{x}, R)$, is minimised in an unconstrained manner, for a given value of R. Minimisation is continually repeated using the final variable vector from the previous optimisation, with decreasing values of R, until the required tolerance is achieved.

6.9.3 Simulation Results

Optimisation of the zero–order PLPs was not possible with this adaption technique, due to the average and peak power constraints. The addition of these constraints reduces the linearity improvement using these predistorters to negligible levels. This effect occurs because of the limited degrees of freedom that these PLP configurations have.

The linearity improvements using this optimisation technique are depicted in Figure 6.28. To allow comparison with the results from the initial adaption scheme, the discontinuous PLP linearity for that scheme is also shown. In all cases it can be seen that the continuous and discontinuous PLPs provide almost identical levels of linearity improvement. Therefore, the continuous PLPs should be used in preference, since they have much reduced complexity compared to the discontinuous PLPs.

The performance of the PLPs has been significantly improved with respect to the initial adaption mechanism. This is particularly evident in the peak distortion improvement. The performance difference between the TC–PLP and the G–PLP for the initial adaption scheme was only slight. Using this optimisation technique this is no longer the case, the TC–PLP has a significantly improved performance.

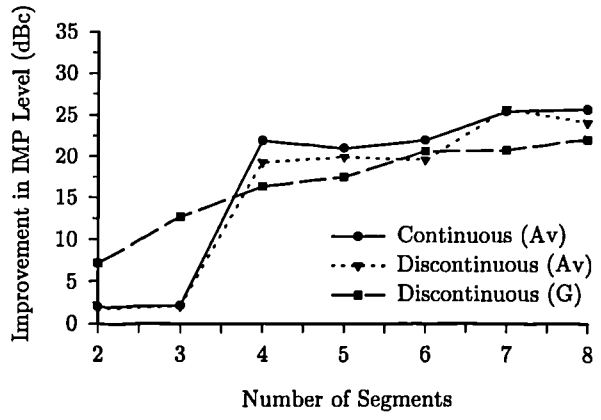
The improvement in linearity evident with the use of this optimisation scheme allows for markedly reduced PLP complexity for a given linearity. For example, to achieve 15dB of average distortion reduction using a continuous TC–PLP requires 7 segments in the equi–spaced case with the initial adaption scheme, and only 3 segments in this case. Since two PLPs are required to operate on the quadrature components, with two control variables for each knot, the number of control elements required in this case is reduced from 28 to only 12.

6.10 Summary

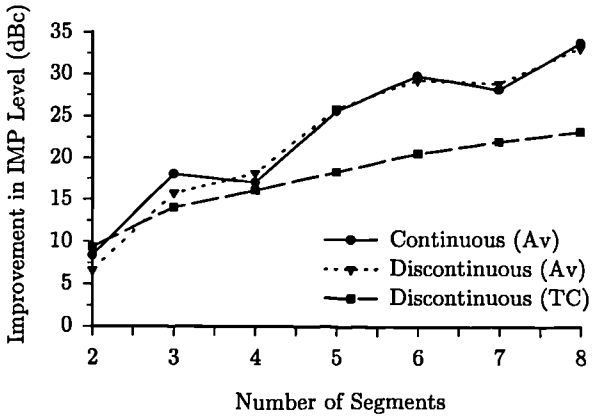
This chapter has examined the piecewise–linear predistortion technique in detail. Several possible architectures, and practical configurations have been suggested. There are two types of PLP, the gain type–PLP (G–PLP) and the transfer characteristic type–PLP (TC–PLP), and three subdivisions of each type (zero–order, continuous first–order and discontinuous first–order). All six of these architectures have been examined in depth, using computer simulation, to ascertain the linearity improvement which each can provide.

In order to optimise the performance of the PLP, two adaption mechanisms have been proposed. The initial adaption scheme relied on minimising an error signal generated by the difference between the actual and ideal output envelopes. In a practical system this would require the real–time analysis of the output signals; this may not be possible in the wideband case. However, it does enable a useful guide to the levels of linearity which can be achieved. Using a 4 segment PLP, it has been determined that the class C amplifier linearity (in terms of maximum IMP) could be improved by more than 16dB. This is a similar level of performance to that attained in the practical class C feedforward amplifier.

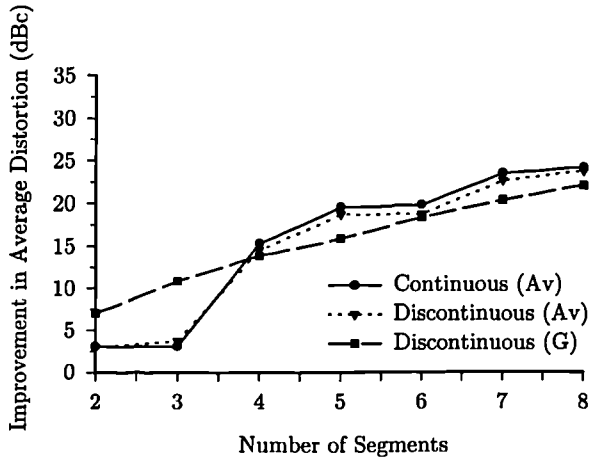
To allow a more realistic comparison of linearity, it has been proposed that the average or peak distortion level should be used as the performance metric. This may be found using the first loop in a feedforward system. Using the initial adaption mechanism, it has been



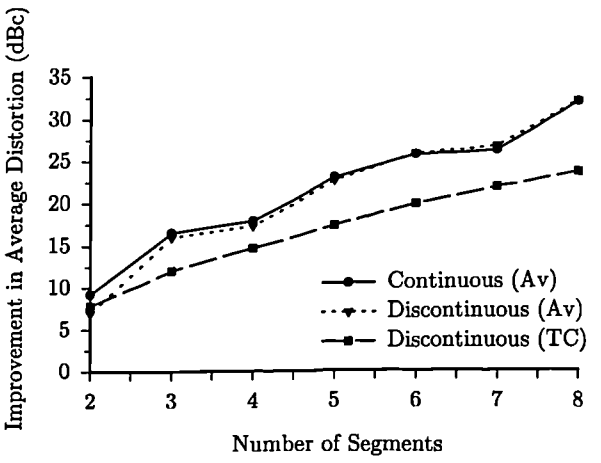
(a) Improvement in highest IMP (G-PLP)



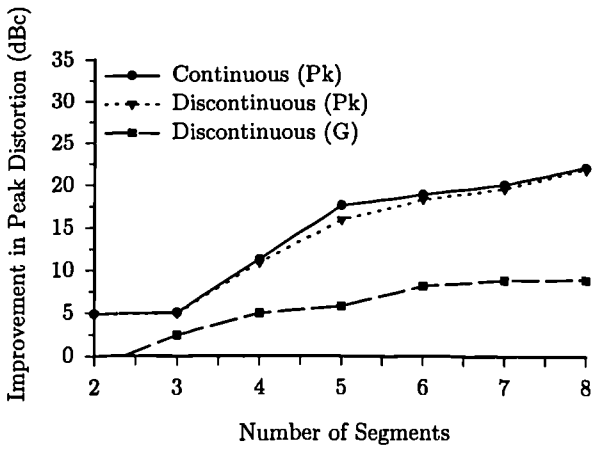
(b) Improvement in highest IMP (TC-PLP)



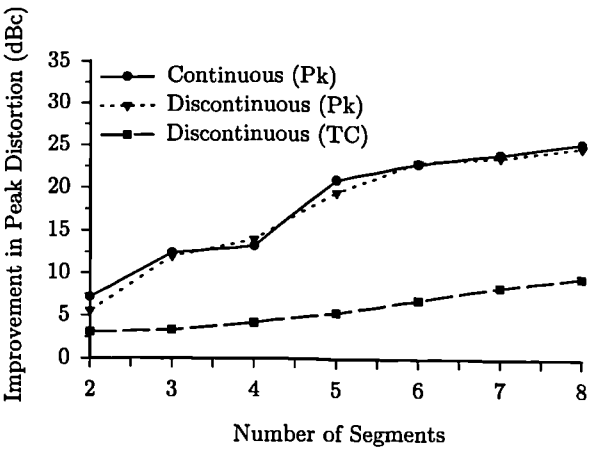
(c) Improvement in average distortion (G-PLP)



(d) Improvement in average distortion (TC-PLP)



(e) Improvement in peak distortion (G-PLP)



(f) Improvement in peak distortion (TC-PLP)

Figure 6.28: Linearity improvement of predistortion amplifier using PLP with optimised characteristic.

shown that the average distortion level is reduced more than the peak distortion level, since no direct account of the distortion level is used.

The initial adaption technique used equally spaced knots in the PLP characteristics; it has

been shown that this is suboptimal. In order to improve performance, a global least–squared fitting procedure with free knots has been used; there was a significant performance increase using this technique. However, since the method cannot take account of time–variant amplifier characteristics, it may not be used in a practical system.

An alternative optimisation scheme has been devised, by considering the adaption as an N–dimensional minimisation problem. Since this method does not rely on the real–time signal envelopes, it may be used in a practical wideband system. It has been found that this technique may not be used with the zero–order PLPs, due to the addition of various constraints on the predistorter knots. It has also been found that the results for the continuous and discontinuous PLPs converge; thus the continuous PLPs should be used in practice, due to their reduced complexity. Using this method, the performance of the PLPs was considerably increased, with the TC–PLP offering improved performance compared with a G–PLP of similar order.

To implement the piecewise–linear predistortion technique in practice requires the design of both the PLP elements and a control mechanism. A number of practical PLP implementations have been suggested. To obtain optimal linearity at all times would require an adaption scheme similar to the N–dimensional minimisation method suggested. This may be implemented using a DSP, since it is required to adapt relatively slowly. However, acceptable performance may be obtained without any adaption, by fixing the control parameters during production, greatly reducing the complexity.

REFERENCES

[1] P. CHARAS AND J.D. ROGERS. Improvements in on–board TWTA performance using amplitude and phase predistortion. In *IEE Int. Conf. on Satellite Communication System Techniques*, volume CP 126, pages 270–280, Birmingham, UK, April 1975.

[2] M.T. ABUELMA'ATTI. *Reduction of intermodulation in communications amplifiers by pre–correction techniques.* PhD thesis, University of Bradford, 1979.

[3] L.L. SCHUMAKER. *Spline functions: basic theory.* J. Wiley & Sons, 1981.

[4] M.T. ABUELMA'ATTI. Synthesis of non–monotonic single–valued function generators without using operational amplifiers. *International Journal of Electronics*, 51(6):803–809, 1981.

[5] V.G. BELLO. Design of a diode function generator using the diode equation and iteration. *IEEE Transactions on Circuit Theory*, CT–19(2):213–214, March 1972.

[6] W.N. CHEUNG. A simplified approach to function generation. *Electronic Engineering*, 54(663):23–27, March 1982.

[7] L.O. CHUA AND F. AYROM. Designing non–linear single op–amp circuits: A cookbook approach. *International Journal on Circuit Theory and Applications*, 13(3):235–268, July 1985.

[8] Hewlett Packard. *Communications Components Designer's Catalog, GaAs and Silicon Products.*

[9] Data Sheet OPA640, Burr–Brown Corporation, 1994.

[10] G.L. BOLTA AND R.A. BRUNS. Zener–diode function generator. *Instruments and Control Systems*, 40(2):135–137, February 1967.

[11] G. KLEIN AND H. HAGENBEUK. Accurate triangle–sine converter. *Electronic Engineering*, pages 700–704, November 1967.

[12] Mini–Circuits. *RF/IF Designer's Handbook*, 1992.

[13] C.J. PAULL AND W.A. EVANS. Waveform shaping techniques for the design of signal sources. *The Radio and Electronic Engineer*, 44(10):523–532, October 1974.

[14] A.R. MANSELL AND A. BATEMAN. Practical implementation issues for adaptive predistortion transmitter linearisation. In *IEE Colloquium on Linear RF Amplifiers and Transmitters*, pages 5/1–5/7, April 1994.

[15] P.D. LOACH AND A.J. WATHEN. On the best least squares approximation of continuous functions using linear splines with free knots. *IMA Journal of Numerical Analysis*, 11:393–409, 1991.

[16] M.J. BAINES. Algorithms for optimal discontinuous piecewise linear and constant L_2 fits to continuous functions with adjustable nodes in one and two dimensions. *Mathematics of Computation*, 62(206):645–669, April 1994.

[17] C.K. CHUI, P.W. SMITH, AND J.D. WARD. On the smoothness of best L_2 approximants from nonlinear spline manifolds. *Mathematics of Computation*, 31:17–23, 1977.

[18] W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY. *Numerical recipes in C - The art of scientific computing.* Cambridge University Press, second edition, 1992.

[19] G.V. REKLAITAS, A. RAVINDRAN, AND K.M. RAGSDELL. *Engineering optimisation – Methods and applications.* J. Wiley & Sons, New York, USA, 1983.

[20] B.D. BUNDAY AND G.R. GARSIDE. *Optimisation methods in Pascal.* Edward Arnold Ltd., London, 1987.

[21] S.L.S. JACOBY, J.S. KOWALIK, AND J.T. PIZZO. *Iterative methods for nonlinear optimisation problems.* Prentice–Hall, New Jersey, USA, 1972.

[22] J.L. KUESTER AND J.H. MIZE. *Optimisation techniques with Fortran.* McGraw–Hill Book Company, New York, USA, 1973.

Chapter Seven

Conclusions and Suggestions for Future Work

7.1 Conclusions

7.1.1 The need for linear amplification

There are two main applications which require the use of a linear amplifier for transmission. Firstly, to improve spectral–efficiency, linear modulation schemes may be used. Secondly, multi–carrier applications can operate with increased flexibility and power efficiency, by summing the channels at low power and then amplifying the resultant signal, rather than performing the addition at high power.

This work relates to the design of an RF amplifier suitable for such applications. In the vast majority of cases, a traditional linear amplifier cannot be used, since to obtain the necessary linearity requires an amplifier with an excessive power rating, operating with very poor efficiency. Therefore, some form of amplifier linearisation technique is required.

The majority of the amplifier linearisation techniques are restricted to narrow channel bandwidths, due to either their use of a digital signal processor (DSP) to generate signals in real–time (in the case of adaptive baseband predistortion); conditional stability (in the case of continuous feedback), or because the signals used in the linearisation are more wideband than the input signal (in the case of LINC or EE&R). These factors prohibit use of these techniques in wideband multi–carrier applications. There are also a number of linear modulation schemes proposed for future systems, in which the modulation bandwidth is too great for these narrowband schemes to be used.

The only linearisation techniques which can be used for broadband operation are feedforward linearisation, and analogue predistortion. Feedforward linearisation operates by generating an error signal which consists almost entirely of distortion. This is then injected into the output signal in antiphase in order to cancel the distortion. Analogue predistortion uses a nonlinear network which (ideally) has the inverse transfer characteristic to the amplifier. Thus, the cascaded system operates in a linear fashion. These techniques have been extensively studied in this work, and new techniques have been presented in order to improve their performance.

7.1.2 Feedforward Linearisation

Feedforward linearisation has been used for some time to produce amplifiers with excellent linearity; however, its power efficiency characteristics have been largely unstudied. New work has been presented to address this question, and techniques have been suggested to improve its efficiency.

The efficiency of a feedforward amplifier is inherently reduced by its requirement for an additional amplifier (the error amplifier) to cancel the distortion. Thus, the efficiency of the error amplifier can be an important determining factor in the overall efficiency, if the error signal is significant (*e.g.*, if a class C main amplifier is used). The other major factors which determine the power efficiency have been found to be the coupling factor of the output coupler used to inject the error signal, and the insertion loss of the main path delay.

A high coupling factor (*e.g.*, 20dB) results in low main path loss, but a large error path loss; requiring a high error amplifier power output. A low coupling factor (*e.g.*, 3dB) reduces the error path loss, but increases the main path loss. Consequently, a compromise between these extremes produces an optimal coupling factor.

The insertion loss of the main path delay will cause a reduction in the output power, and thus lowers the overall efficiency (although a lower error amplifier power will be required).

This reduction in efficiency can be highly significant, and so there is much to be gained by reducing the delay insertion loss. One method by which this may be achieved is to reduce the delay time, which will have implications on the distortion cancellation which can be attained. The use of the variable phase shift in the error path will allow the two cancellation paths to be phase matched at one frequency (usually chosen to be the centre of the band), whilst the cancellation will reduce further from this frequency, since the two paths are not time matched. The degree of cancellation which can be attained across the band has been theoretically analysed and shown to be dependent upon the relative bandwidth. In many communications systems the relative bandwidth is small, and so this technique may provide adequate linearity, whilst significantly improving efficiency.

A practical feedforward system has been constructed, operating at a centre frequency of 220MHz, in order to verify the theoretical efficiency analysis. The system has been designed with the priority being to maintain efficiency; thus class C main and error amplifiers have been used. Although the use of a nonlinear error amplifier limits the linearity improvement which the feedforward technique can provide, the overall linearity was still reasonable, commensurate with that of a class A amplifier (in terms of highest intermodulation distortion product (IMP)). The overall efficiency achieved in practice was close to that predicted, and was significantly higher than that which could be attained by a conventional amplifier with similar linearity. The main path delay was removed to investigate the effect of reducing the delay. It was found that the efficiency was increased, without a significant degradation in the linearity.

An alternative technique which offers an improvement in the efficiency of a feedforward amplifier is to recover some of the power lost in the feedforward process. The majority of the power generated by the error amplifier is wasted in the output coupler termination. It is proposed that by replacing the coupler termination with a matched rectifier, this power may be recovered, thus improving overall efficiency. However, this technique can only provide a significant increase in efficiency if the error amplifier power is a substantial fraction of the main amplifier power, *i.e.*, the main amplifier is highly nonlinear. This technique may be of use in satellite systems, since the linearity requirement is not so demanding, whilst any reduction in power consumption is of considerable benefit.

Whilst the previous two techniques for improving the efficiency of a feedforward amplifier are useful, the efficiency is still degraded by the application of feedforward, due to the error amplifier power. A potential method to alleviate this is to use analogue predistortion of the main amplifier. In such a system the required error amplifier power will be lowered, resulting in improved efficiency. The overall efficiency will also be much less sensitive to the error amplifier efficiency, allowing a more linear error amplifier to be used, thereby improving the overall linearity. A practical system has been constructed, operating at a centre frequency of 850MHz, to verify the benefits of the combined use of feedforward and predistortion.

The analogue predistortion techniques which are in general use are crude, and only offer a small improvement in linearity. It would also appear that analogue predistortion of class C amplifiers has not been studied. Therefore, analogue predistortion has been further investigated.

7.1.3 Broadband Predistortion Linearisation

Broadband predistortion linearisation is generally implemented using an analogue nonlinear network operating at either RF or IF. There are a number of possible nonlinear elements and architectures which may be used, several of which have been described.

The nonlinear elements which may be used for analogue predistortion fall into two categories. The polynomial predistorter uses a separate nonlinearity for each order of distortion to be reduced. This is not applicable for predistortion of highly nonlinear amplifiers, since a large number of networks would be required, making the technique prohibitively complex. In contrast, the generic predistorter uses a single nonlinear network, which generates many orders of distortion, and so allows highly nonlinear amplifiers to be predistorted.

The nonlinear networks used in predistortion may be arranged in one of several architectures. The scalar predistorter uses one nonlinear network to reduce the AM/AM or AM/PM conversion of the amplifier; it is unable to simultaneously reduce both. The vector predistorter uses two nonlinear networks in polar or quadrature form, and allows simultaneous reduction of both the AM/AM and AM/PM conversion. The quadrature form is most commonly used, due to its symmetrical nature.

The predistortion technique which offers the best linearity performance is that of the generic vector predistorter, and this is used as the basis for further study.

The characteristic of an ideal predistorter has been derived, and shown to be the inverse of the amplifier characteristic. To allow the ideal characteristic to be derived from measured amplifier data, the AM/AM and AM/PM conversion characteristics of an ideal predistorter have also been derived from those of the amplifier. A computer simulation has been described, which allows modelling of a predistortion amplifier. The amplifier model utilises the measured transfer characteristic of the class C main amplifier used in the 220MHz feedforward system previously described; the ideal predistortion characteristic is derived from this measured data.

Using this computer simulation, with a two–tone input signal, it has been found that the amplifier model produces a good approximation to the practical amplifier. The envelope of the signal, and the two–tone output spectrum correspond well with those measured. The envelope of the ideal predistorter output has been found, and this is used in later work to allow comparison of non–ideal predistorter characteristics. The output of the predistortion amplifier has been shown to contain some IMPs, but at a negligible level (below -120dBc). These are caused by the fact that the amplifier data is only measured at a finite number of points, requiring interpolation between them for both the amplifier and the predistorter models. Since the level of the IMPs is so low, this does not invalidate the ideal predistorter characteristic generated.

The instantaneous power efficiency of an ideal predistortion amplifier has been derived. It has been shown that if the power dissipated by the predistorter network is negligible compared with that of the amplifier, the instantaneous efficiency referenced to the output is unaffected by predistortion. The instantaneous efficiency referenced to the input is, however, affected since the predistorter alters the statistics of the amplifier input signal.

The effect of nonlinear amplification and predistortion have been further investigated by the use of probability density functions (p.d.f.s). To enable this analysis the output signal p.d.f. of a nonlinearity with a given input signal p.d.f. has been derived. A number of representative signals have been applied to the amplifier and predistorter in order to illustrate the nonlinear effects. Using the class C amplifier model, it has been found that the amplifier increases the probability of the output signal being close to the maximum or minimum amplitudes. This is due to the switch–like nature of the amplifier characteristic. In order to compensate for this effect, the predistorter enhances the probability that its output signal will be in the middle of the range, and reduces the probability that it will be at the extremes.

The average efficiency of a predistortion amplifier has been calculated using the representative input signals and measured characteristics from class A and C amplifiers. The average efficiency of the predistortion amplifier is lower than that of the amplifier alone, due to the

modification of the input signal p.d.f.. It has been shown that predistortion of a class C amplifier has a more detrimental effect on the efficiency than for a class A amplifier. This effect is due to fact that the transfer characteristic of the class A amplifier is more linear, and hence not affected by predistortion to the same degree as the class C amplifier. The reduction in average efficiency caused by the application of predistortion has also been shown to be more significant for signals with high peak–to–mean ratios. The average efficiency has been compared with that of the practical feedforward system, and it is found that (in at least the case discussed), feedforward linearisation degrades efficiency to a far greater extent than predistortion.

7.1.4 Piecewise–Linear Predistortion

The linearity improvement offered by current forms of analogue predistortion is limited. This has thus far constrained the use of analogue predistortion to amplifiers which are only mildly nonlinear, such as saturating class AB, and travelling wave tube amplifiers (TWTAs). To improve power efficiency, class C amplifiers could be used, however, these are highly nonlinear, and so are more difficult to linearise effectively using predistortion. In contrast with mildly nonlinear amplifiers, class C amplifiers produce increased distortion (relative to the linear part) as the amplifier is backed–off. Thus, it is not possible to back the amplifier off to meet the required linearity specification.

An improved method of analogue predistortion, using a piecewise–linear predistorter (PLP), has been investigated. This technique approximates the required predistorter characteristic using a number of piecewise–linear segments, and offers the potential of improved approximation compared with other methods. A number of possible architectures have been described, and a computer simulation is used to ascertain their linearity improvement.

There are two basic forms of PLP which may be used. The gain–type PLP (G–PLP) approximates the complex gain of the ideal predistorter, whilst the transfer characteristic–type PLP (TC–PLP) approximates the ideal predistorter transfer characteristic. Each of these forms may be subdivided into three configurations, which are dependent upon placing constraints on the knots of the predistorter. The zero–order PLP approximates the relevant characteristic using piecewise–constant segments, and is the simplest configuration. The continuous first–order PLP approximates the relevant characteristic using piecewise–linear segments, forming a characteristic which is constrained to be continuous. The discontinuous first–order PLP is similar to the continuous first–order PLP, but does not have the constraint of continuity.

In order to optimise the predistorter architectures, some form of adaption mechanism is required. An adaption technique similar to that used in baseband predistortion was initially chosen. The method uses an error signal determined by the difference of the actual predistortion amplifier output to its ideal. This error signal is then minimised by a linear adaption process in order to optimise performance. This technique is not suitable for use in a practical broadband predistortion system, since the error signal is derived in real–time from the time–varying signal envelopes. However, its use in simulation is acceptable, and provides an indication of the linearity which can be attained.

The standard performance metric used to provide an indication of linearity is to measure the highest IMP level. This is not a particularly useful metric in this application, since the levels of IMPs after linearisation are still significant, and occupy a wider bandwidth than the output of the amplifier alone. An alternative performance metric has been proposed, which is derived by measuring the amount of distortion generated by the amplifier, either as a peak or an average value. This signal is determined using the first loop in a feedforward system to

cancel those frequencies present in the input signal, leaving just the distortion.

The simulation results obtained using the initial adaption technique show that the PLP significantly improves the linearity of the class C amplifier. Using a 4 segment discontinuous PLP it is possible to reduce the highest IMP level by greater than 16dB.

To improve the linearity which can be obtained using the PLP, it is necessary to alter the knot spacing. This could not be done efficiently using the initial adaption method, and so an alternative has been used. The method chosen used a global least–squared fitting procedure in order to find the best complex gain or transfer characteristic for the PLP. Using this procedure, the performance of the continuous TC–PLP was significantly improved. The linearity of the optimal least–squared fit continuous TC–PLP was improved by up to 5dB, compared with the equi–spaced discontinuous TC–PLP, in terms of highest IMP and average distortion, and by up to 10dB in terms of peak distortion. The performance of the optimal least–squared fit continuous G–PLP was generally not as good as the equi–spaced discontinuous G–PLP, in terms of highest IMP and average distortion, although its peak distortion was significantly improved.

Using the alternative performance metrics, another optimisation method has been derived, which directly minimises the average or peak distortion. This method is based on standard N–dimensional minimisation schemes, with the addition of constraints to avoid knot entanglement, and ensure constant gain through the predistortion amplifier. This is a more practical method than the initial technique, since the error signal is not found in real–time. This technique is particularly suitable for this application, since the predistortion linearisation is to be used to supplement feedforward linearisation. Thus the distortion signal will already be present in the system. It should be noted that the optimisation scheme will only converge to a local minimum, rather than the global one, and thus, whilst the results presented illustrate that significant improvements in linearity result from the use of the PLP, it may be that even better results may be obtained. The increased linearity attained using this optimisation method allows the complexity of the PLP to be significantly reduced, for a given level of linearity.

7.2 Future Work

7.2.1 Feedforward Linearisation

An improved method to determine the efficiency of a feedforward amplifier can be developed, using similar methods to those used to determine the average efficiency of a predistortion amplifier. This would enable more accurate average efficiency modelling in a feedforward system with highly nonlinear amplifiers, since all orders of IMP can be taken into account. This is not the case with the theoretical analysis presented in this work, as this assumed only third–order distortion. This would enable a more rigorous comparison of the average efficiency of predistortion and feedforward linearisation.

It would be instructive to construct a practical feedforward system using the power recovery technique, in order to validate the theoretical analysis. In satellite systems, the extra complexity of this technique may be justified, since the linearity improvement is not required to be great, but efficiency must be maintained.

The use of feedforward linearisation has thus far been restricted to wideband systems, with other techniques being used in preference for narrowband applications. This is due to the perceived increased complexity and reduced power efficiency when compared with these nar-

rowband methods. However, this work has shown that this need not be the case, since in systems with small relative bandwidth, the time delays may be removed without impairing linearity, whilst dramatically increasing efficiency. The performance may be improved further with the use of analogue predistortion of the main amplifier. Further investigation is required to determine whether an adaptive feedforward system can compare favourably in the narrowband case with other linearisation techniques.

7.2.2 Piecewise–Linear Predistortion

The most important work still required with the piecewise–linear predistortion technique is to construct a practical system. This would enable a comparison to be made of the linearity of the practical and the simulated systems. The increased degrees of freedom which the technique offers will require a sophisticated optimisation scheme to be implemented in practice. The methods required to obtain the performance metrics used to optimise the system have already been investigated for their use in feedforward linearisation, and so should be simple to obtain.

The models used in the simulation of the predistortion amplifier were memoryless, and hence do not include the effects of frequency–dependence of the amplifier and predistorter characteristics, or of IMP asymmetry. The former of these could be implemented by using a simulation with increased complexity. However, IMP asymmetry is a relatively unstudied phenomenon, and so no models exist which take it into account. Therefore, an investigation into the causes of asymmetry is required, with the aim to reduce its effect in practical systems, and allow its modelling. These effects are likely to impair the performance of the analogue predistortion, and so should be investigated.

It is possible for the PLP to operate on the instantaneous RF signal (or an IF equivalent) rather than the envelope. This will allow for more broadband operation, since an envelope detector is no longer required (the output of which is generally more broadband than the modulation signal). However, since the majority of practical circuits are unipolar, double the number of piecewise–linear generating elements (usually a resistor–diode combination) and control variables would be required. This configuration may not be directly simulated, since the instantaneous nonlinearity will generate significant harmonics, causing aliasing. However, an envelope equivalent configuration, similar to that of the continuous transfer characteristic–type PLP (TC–PLP) may be used. This requires the envelope equivalent transfer characteristic of an instantaneous PLP to be derived, which may be obtained using the first–order Chebyshev Transform. It would be expected that the results thus obtained would be similar to those of the continuous TC–PLP, due to the similarity of the configurations.

APPENDIX A

COMPLEX ENVELOPE EQUIVALENT OF A MEMORYLESS NONLINEARITY

The complex envelope models of the amplitude and phase and quadrature representations of a bandpass nonlinearity are shown in Figure 2.6. This Appendix will show that the complex envelope and standard representations are equivalent.

A.1 Amplitude and Phase Representation

If a general bandpass input signal, $x(t)$, consisting of a single modulated carrier given by

$$x(t) = A(t)\cos(\omega t + \phi(t)) \tag{A.1}$$

is applied to a nonlinearity in amplitude and phase form, the output signal, $y(t)$, has been shown in Section 2.4.1 to be given by (making time implicit for the modulation)

$$y(t) = F(A)\cos(\omega t + \phi + G(A)) \tag{A.2}$$

Consider the input signal, $x(t)$, given in analytical form, thus

$$x(t) = Ae^{\jmath(\omega t+\phi)} \tag{A.3}$$

The output of the complex envelope form of the amplitude and phase representation (Figure 2.6(a)) is given by

$$y(t) = F(A)e^{\jmath G(A)}\left(\frac{x(t)}{|x(t)|}\right) \tag{A.4}$$

It may be simply shown that

$$\frac{x(t)}{|x(t)|} = e^{\jmath(\omega t+\phi)} \tag{A.5}$$

Therefore,

$$y(t) = F(A)e^{\jmath G(A)}e^{\jmath(\omega t+\phi)} \tag{A.6}$$

Or

$$y(t) = F(A)e^{\jmath(\omega t+\phi+G(A))} \tag{A.7}$$

This is equivalent to

$$y(t) = F(A)(\cos(\omega t + \phi + G(A)) + \jmath \sin(\omega t + \phi + G(A))) \tag{A.8}$$

Therefore, the real part of the output signal, $\Re[y(t)]$, is given by

$$\Re[y(t)] = F(A)\cos(\omega t + \phi + G(A)) \tag{A.9}$$

Which is equivalent to the output of a memoryless bandpass nonlinear device in amplitude and phase form, given in Equation A.2.

A.2 Quadrature Representation

If a general bandpass input signal, $x(t)$, consisting of a single modulated carrier given by

$$x(t) = A(t)\cos(\omega t + \phi(t)) \tag{A.10}$$

is applied to a nonlinearity in quadrature form, the output signal, $y(t)$, has been shown in Section 2.4.2 to be given by (making time implicit for the modulation)

$$y(t) = P(A)\cos(\omega t + \phi) - Q(A)\sin(\omega t + \phi) \tag{A.11}$$

Consider the input signal, $x(t)$, given in analytical form, thus

$$x(t) = Ae^{\jmath(\omega t+\phi)} \tag{A.12}$$

The output of the complex envelope form of the quadrature representation (Figure 2.6(b)) is given by

$$y(t) = (P(A) + \jmath Q(A))\left(\frac{x(t)}{|x(t)|}\right) \tag{A.13}$$

It may be simply shown that

$$\frac{x(t)}{|x(t)|} = \cos(\omega t + \phi) + \jmath \sin(\omega t + \phi) \tag{A.14}$$

Hence

$$\begin{aligned} y(t) = {} & (P(A)\cos(\omega t + \phi) - Q(A)\sin(\omega t + \phi)) \\ & + \jmath(P(A)\sin(\omega t + \phi) + Q(A)\cos(\omega t + \phi)) \end{aligned} \tag{A.15}$$

Therefore, the real part of the output signal, $\Re[y(t)]$, is given by

$$\Re[y(t)] = P(A)\cos(\omega t + \phi) - Q(A)\sin(\omega t + \phi) \tag{A.16}$$

Which is equivalent to the output of a memoryless bandpass nonlinear device in quadrature form, given in Equation A.11.

Appendix B

Theoretical Analysis of Feedforward Linearisation

B.1 Analysis of Feedforward Amplifier Linearity[1]

Referring to Figure B.1, assume that the input, $x(t)$, is a single CW tone of amplitude A, thus

$$x(t) = A\cos(\omega t + \phi) \tag{B.1}$$

Therefore, the output from C2, $y(t)$, is given by

$$y(t) = A\cos\big(\omega(t-\tau_1)+\phi\big) + \hat{A}\cos\big(\omega(t-\tau_2)+\phi+\hat{\theta}\big) \tag{B.2}$$

where $\hat{A}$ is the amplitude of the error path signal, and $\hat{\theta}$ is the phase difference between the paths.

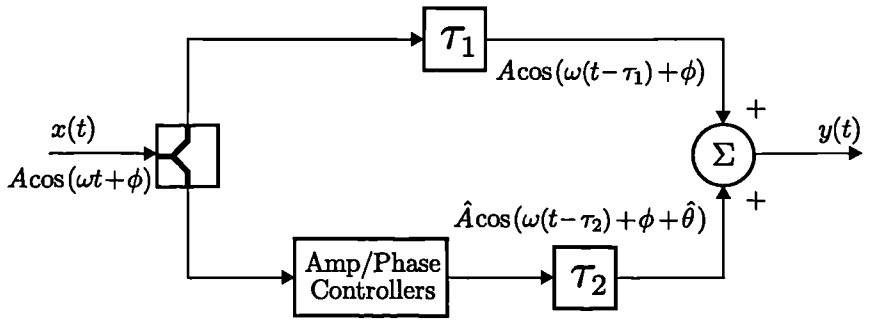


Figure B.1: Simplified configuration of a single loop in a feedforward amplifier.

Assuming that the difference in delay between the two paths is $\Delta\tau$, the error path delay is given by

$$\tau_2 = \tau_1 + \Delta\tau \tag{B.3}$$

[1]The analysis presented in this section is an extended version of that published by the author in [1]

This difference in the time delays is termed *delay mismatch* [1]. Substitution of Equation B.3 into Equation B.2 gives

$$y(t) = A\cos\big(\omega(t-\tau_1)+\phi\big) + \hat{A}\cos\big(\omega(t-\tau_1-\Delta\tau)+\phi+\hat{\theta}\big) \tag{B.4}$$

Expanding gives

$$y(t) = A\cos\left(\omega(t-\tau_1)+\phi\right)\cdot\mathcal{P} + A\sin\left(\omega(t-\tau_1)+\phi\right)\cdot\mathcal{Q} \tag{B.5}$$

where

$$\mathcal{P} = 1+\left(\frac{\hat{A}}{A}\right)\cos\left(\omega\Delta\tau-\hat{\theta}\right) \tag{B.6}$$

$$\mathcal{Q} = \left(\frac{\hat{A}}{A}\right)\sin\left(\omega\Delta\tau-\hat{\theta}\right) \tag{B.7}$$

The resultant output signal will consist of a single tone with phase γ, nominally delayed by τ_1, hence

$$y(t) = R\cos\left(\omega(t-\tau_1)+\gamma\right) \tag{B.8}$$

Thus

$$R\cos\left(\omega(t-\tau_1)+\gamma\right) = A\cos\left(\omega(t-\tau_1)+\phi\right)\cdot\mathcal{P} + A\sin\left(\omega(t-\tau_1)+\phi\right)\cdot\mathcal{Q} \tag{B.9}$$

The amplitude of the output signal, R, can be found from

$$R^2 = (A\mathcal{P})^2 + (A\mathcal{Q})^2 \tag{B.10}$$

Substituting for $\mathcal{P}$ and $\mathcal{Q}$ yields

$$R^2 = A^2\left[1+\left(\frac{\hat{A}}{A}\right)\cos\left(\omega\Delta\tau-\hat{\theta}\right)\right]^2 + A^2\left[\left(\frac{\hat{A}}{A}\right)\sin\left(\omega\Delta\tau-\hat{\theta}\right)\right]^2 \tag{B.11}$$

Simplification gives

$$R^2 = A^2 + \hat{A}^2 + 2A\hat{A}\cos\left(\omega\Delta\tau-\hat{\theta}\right) \tag{B.12}$$

Thus, the amplitude of the output signal, R, is given by

$$R = \sqrt{A^2+\hat{A}^2+2A\hat{A}\cos\left(\omega\Delta\tau-\hat{\theta}\right)} \tag{B.13}$$

The cancellation (in decibels), C_{dB}, of the signals in the two paths, relative to A, is

$$C_{dB} = 20\log_{10}(A) - 20\log_{10}(R) \tag{B.14}$$

$$= -10\log_{10}\left\{1+\left(\frac{\hat{A}}{A}\right)^2 + 2\left(\frac{\hat{A}}{A}\right)\cos\left(\omega\Delta\tau-\hat{\theta}\right)\right\} \tag{B.15}$$

To find the phase difference, $\Delta\gamma$, between the input tone and the output tone, the output phase, γ, is referenced to the input phase, ϕ, yielding

$$\Delta\gamma = \tan^{-1}\left(\frac{A\mathcal{Q}}{A\mathcal{P}}\right) \tag{B.16}$$

Substituting for $\mathcal{P}$ and $\mathcal{Q}$ gives

$$\Delta\gamma = \tan^{-1}\left(\frac{\left(\frac{\hat{A}}{A}\right)\sin\left(\omega\Delta\tau - \hat{\theta}\right)}{1 + \left(\frac{\hat{A}}{A}\right)\cos\left(\omega\Delta\tau - \hat{\theta}\right)}\right) \tag{B.17}$$

To achieve optimal cancellation would require that the amplitudes of the two paths be equal, and their phase difference be π, at all frequencies. Thus, the delay mismatch is required to be zero. Hence,

$$\hat{A} = A \tag{B.18}$$

$$\hat{\theta} = \pi \tag{B.19}$$

$$\omega\Delta\tau = 0 \tag{B.20}$$

In practice, the actual amplitude and phase difference ($\hat{A}$ and $\hat{\theta}$) will differ from the ideal values. The amplitude error is denoted by δA, and phase difference between the paths is given by $\theta + \delta\theta$, where θ is the gross phase shift, and $\delta\theta$ takes account of any further phase error, thus

$$\hat{A} = A + \Delta A \tag{B.21}$$

$$\hat{\theta} = \theta + \delta\theta \tag{B.22}$$

Equation B.21 may be normalised to A to give

$$\left(\frac{\hat{A}}{A}\right) = 1 + \delta A \tag{B.23}$$

With these values the cancellation and phase shift expressions (Equations B.15 and B.17) are given by

$$C_{dB} = -10\log_{10}\left\{1 + (1+\delta A)^2 + 2(1+\delta A)\cos\left(\omega\Delta\tau - \theta - \delta\theta\right)\right\} \tag{B.24}$$

$$\Delta\gamma = \tan^{-1}\left(\frac{(1+\delta A)\sin\left(\omega\Delta\tau - \theta - \delta\theta\right)}{1 + (1+\delta A)\cos\left(\omega\Delta\tau - \theta - \delta\theta\right)}\right) \tag{B.25}$$

In general, the delay mismatch will not be zero, and so the cancellation becomes frequency–dependent. To achieve perfect cancellation at a frequency, ω_0, (usually the centre of the frequency band of operation), requires that

$$\omega_0\Delta\tau - \hat{\theta} = n\pi \tag{B.26}$$

Hence, in practice, the gross phase shift will be

$$\theta = \omega_0\Delta\tau - \pi - \delta\theta \tag{B.27}$$

since the phase controller will generally have a phase variation of at most $\pm\pi$.

Intuitively, it would be expected that the amplitudes of the two paths be equal for optimal cancellation. It has been shown that this is not the case if the average cancellation across the operation frequency band is considered [2]. However, the difference in amplitude between the paths is negligible for this application, and so will not be considered.

B.2 Analysis of Standard Feedforward Amplifier Efficiency[2]

B.2.1 System Parameters

Referring to the block diagram for a feedforward amplifier shown in Figure B.2, the parameters used in the analysis are explained below.

Main Amplifier, A1
Output power (per carrier): P_{A1} (W)
Efficiency: η_{A1}
Third–order IMP level: S_{A1} (dBc)

Error Amplifier, A2
Output power (per carrier): P_{A2} (W)
Efficiency: η_{A2}

Output Coupler, C2
Coupling factor: C_f (dB)

Main Path Delay, TD2
Delay Insertion Loss: L_{TD2} (dB)

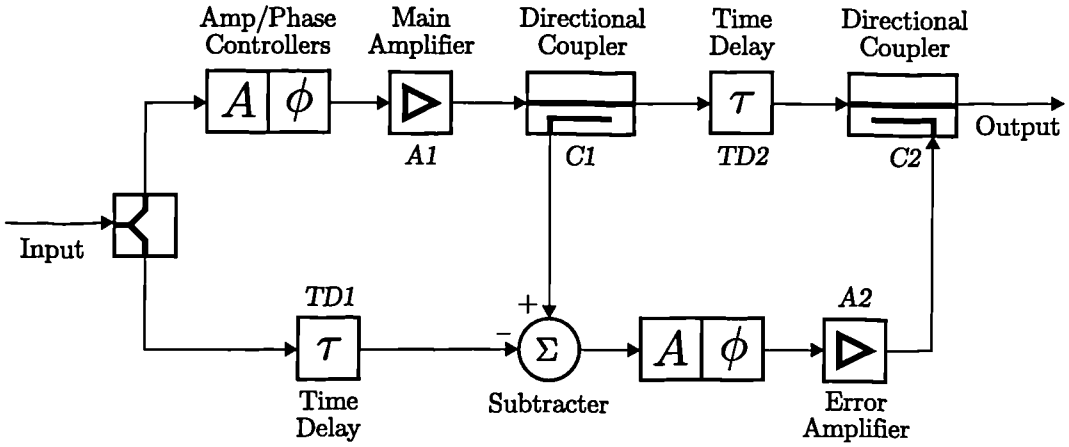


Figure B.2: Configuration of a feedforward amplifier.

B.2.2 Theoretical Analysis

The analysis assumes that the main amplifier generates only third–order IMD, which is completely cancelled across the frequency band, and that the error amplifier produces no distortion. Whilst these assumptions are only an approximation to the practical situation, they allow important trends to be identified.

With a two–tone test signal applied, the fractional power of a single IMP, F_{IM}, (per carrier) is given by

$$F_{IM} = 10^{(S_{A1}/10)} \tag{B.28}$$

[2]The analysis presented in this section is an extended version of that published by the author in [3]

Thus, the absolute power of a single third–order IMP, P_{IM}, is

$$P_{IM} = P_{A1}F_{IM} \tag{B.29}$$

The fractional loss of signals through the delay, L, is

$$L = 10^{(L_{TD2}/10)} \tag{B.30}$$

Thus, both the fundamental signal and the IMPs are attenuated by this factor. Hence

$$P_{A1,L} = \frac{P_{A1}}{L} \tag{B.31}$$

and

$$P_{IM,L} = \frac{P_{A1}F_{IM}}{L} \tag{B.32}$$

The coupling factor of the output coupler C2, C_{DC}, is

$$C_{DC} = 10^{-(C_f/10)} \tag{B.33}$$

Thus, the effective power of the third–order IMP from amplifier A1 is

$$P_{IM,C} = \frac{P_{A1}F_{IM}L_{DC}}{L} \tag{B.34}$$

where L_{DC} is the through loss of the output coupler C2. To determine the through loss of the coupler, assume an incident power, P_I, applied to the coupler. The coupled power, P_C, will be dependent upon the coupling factor, C_{DC}, thus

$$P_C = P_I C_{DC} \tag{B.35}$$

The through path power at the output of the coupler, P_{TP}, assuming no additional losses, is given by

$$P_{TP} = P_I - P_C \tag{B.36}$$

$$= P_I(1 - C_{DC}) \tag{B.37}$$

The through loss of the coupler is given by the ratio of the through path power P_{TP} to the incident power P_I, thus

$$L_{DC} = \frac{P_{TP}}{P_I} = 1 - C_{DC} \tag{B.38}$$

Hence, substituting this into Equation B.34 gives

$$P_{IM,C} = \frac{P_{A1}F_{IM}(1 - C_{DC})}{L} \tag{B.39}$$

The effective power of the IMP from A2 is

$$P_{A2,C} = P_{A2}C_{DC} \tag{B.40}$$

To completely cancel the IMP, $P_{IM,C}$ and $P_{A2,C}$ must be equal and applied in antiphase. Thus

$$P_{A2} = \frac{P_{A1}F_{IM}(1 - C_{DC})}{LC_{DC}} \tag{B.41}$$

The RF output power, $P_{A1,C}$, is given by

$$P_{A1,C} = P_{A1,L}(1 - C_{DC}) \tag{B.42}$$

By substitution of Equation B.31 this becomes

$$P_{A1,C} = \frac{P_{A1}(1 - C_{DC})}{L} \tag{B.43}$$

The overall efficiency of a feedforward amplifier is

$$\eta_{ff} = \frac{P_{A1,C}}{P_{dc,A1} + P_{dc,A2}} \tag{B.44}$$

where $P_{dc,A1}$ and $P_{dc,A2}$ are the dc powers supplied to the amplifiers A1 and A2 respectively and are given by

$$P_{dc,A1} = \frac{P_{A1}(1 + F_{IM})}{\eta_{A1}} \tag{B.45}$$

and

$$P_{dc,A2} = \frac{P_{A2}}{\eta_{A2}} \tag{B.46}$$

It should be noted that Equation B.45 is slightly different to the same quantity defined in [4] as it also includes the effect of the IMP on the DC power supplied to the main amplifier, A1. Substituting for P_{A2} from Equation B.41, Equation B.46 becomes

$$P_{dc,A2} = \frac{P_{A1}F_{IM}(1 - C_{DC})}{\eta_{A2}LC_{DC}} \tag{B.47}$$

Therefore, the overall efficiency is

$$\eta_{ff} = \frac{\eta_{A1}\eta_{A2}C_{DC}(1 - C_{DC})}{\eta_{A1}F_{IM}(1 - C_{DC}) + \eta_{A2}LC_{DC}(1 + F_{IM})} \tag{B.48}$$

Optimal coupling factor

The optimum coupling factor, $C_{DC,OPT}$, can be determined by differentiating Equation B.48, equating to zero and solving for C_{DC} to find the local maximum. Thus

$$\frac{\partial \eta_{ff}}{\partial C_{DC}} = 0 \tag{B.49}$$

Therefore,

$$C_{DC}{}^2(\eta_{A1}F_{IM} - \eta_{A2}L(1 + F_{IM})) - 2C_{DC}\eta_{A1}F_{IM} + \eta_{A1}F_{IM} = 0 \tag{B.50}$$

This can be solved for C_{DC} to find the optimum coupling factor, $C_{DC,OPT}$. Hence

$$C_{DC,OPT} = \frac{\eta_{A1}F_{IM} \pm \sqrt{\eta_{A1}\eta_{A2}F_{IM}L(1 + F_{IM})}}{\eta_{A1}F_{IM} - \eta_{A2}L(1 + F_{IM})} \tag{B.51}$$

However, only the solution using the negative squareroot term leads to a practical coupling factor.

Sensitivity Analysis

The sensitivity of the feedforward amplifier efficiency to the error amplifier efficiency, $S_{\eta_{A2}}^{\eta_{ff}}$, is given by

$$S_{\eta_{A2}}^{\eta_{ff}} = \frac{\partial \eta_{ff}}{\partial \eta_{A2}} \frac{\eta_{A2}}{\eta_{ff}} \tag{B.52}$$

where

$$\begin{aligned} \frac{\partial \eta_{ff}}{\partial \eta_{A2}} = & \frac{\eta_{A1} C_{DC}(1 - C_{DC})}{\eta_{A1} F_{IM}(1 - C_{DC}) + \eta_{A2} L C_{DC}(1 + F_{IM})} \\ & - \frac{\eta_{A1} \eta_{A2} L {C_{DC}}^2 (1 - C_{DC})(1 + F_{IM})}{\left(\eta_{A1} F_{IM}(1 - C_{DC}) + \eta_{A2} L C_{DC}(1 + F_{IM})\right)^2} \end{aligned} \tag{B.53}$$

This simplifies to

$$\frac{\partial \eta_{ff}}{\partial \eta_{A2}} = \frac{{\eta_{A1}}^2 F_{IM} C_{DC}(C_{DC} - 1)^2}{\left(\eta_{A1} F_{IM}(1 - C_{DC}) + \eta_{A2} L C_{DC}(1 + F_{IM})\right)^2} \tag{B.54}$$

Thus, $S_{\eta_{A2}}^{\eta_{ff}}$ is given by

$$S_{\eta_{A2}}^{\eta_{ff}} = \frac{{\eta_{A1}}^2 F_{IM} C_{DC}(C_{DC} - 1)^2}{\left(\eta_{A1} F_{IM}(1 - C_{DC}) + \eta_{A2} L C_{DC}(1 + F_{IM})\right)^2} \times \frac{\eta_{A1} F_{IM}(1 - C_{DC}) + \eta_{A2} L C_{DC}(1 + F_{IM})}{\eta_{A1} C_{DC}(1 - C_{DC})} \tag{B.55}$$

This can be shown to simplify to

$$S_{\eta_{A2}}^{\eta_{ff}} = \frac{\eta_{A1} F_{IM}(1 - C_{DC})}{\eta_{A1} F_{IM}(1 - C_{DC}) + \eta_{A2} L C_{DC}(1 + F_{IM})} \tag{B.56}$$

B.3 Analysis of Feedforward Amplifier Efficiency with Power Recovery

This analysis is based on that presented in Section B.2, and is subject to the same assumptions. Referring to Figure B.3 the power normally lost in output coupler termination, P_{CT}, is now utilised by the power recovery circuitry, and is given by

$$P_{CT} = P_{A2}(1 - C_{DC}) \tag{B.57}$$

This power is utilised with an efficiency of η_{PR}, typically 70–85%, to give a useful power of

$$P_{PR} = \eta_{PR} P_{CT} \tag{B.58}$$

Or

$$P_{PR} = \eta_{PR} P_{A2}(1 - C_{DC}) \tag{B.59}$$

By substitution of Equation B.41 (page 149) this becomes

$$P_{PR} = \frac{\eta_{PR} P_{A1} F_{IM}(1 - C_{DC})^2}{L C_{DC}} \tag{B.60}$$

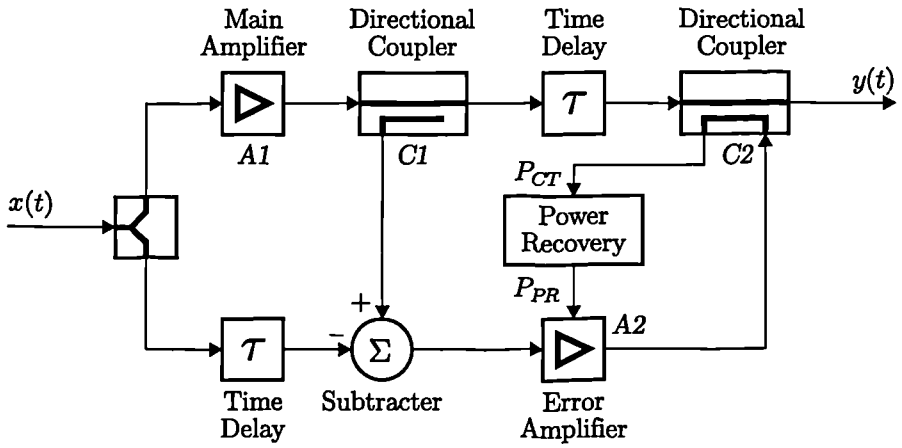


Figure B.3: The configuration of a feedforward amplifier with power recovery.

The overall efficiency of the feedforward amplifier with power recovery is given by

$$\eta_{ff} = \frac{P_{A1,C}}{P_{dc,A1} + P_{dc,A2} - P_{PR}} \tag{B.61}$$

where $P_{dc,A1}$ and $P_{dc,A2}$ are given by Equations B.45 and B.47 respectively. By relevant substitution and simplification this becomes

$$\eta_{ff} = \frac{\eta_{A1}\eta_{A2}C_{DC}(1 - C_{DC})}{\eta_{A1}F_{IM}(1 - C_{DC}) + \eta_{A2}LC_{DC}(1 + F_{IM}) - \eta_{A1}\eta_{A2}\eta_{PR}F_{IM}(1 - C_{DC})^2} \tag{B.62}$$

The optimal coupling factor can be found in a similar manner to that used in Section B.2, and is given by

$$C_{DC,OPT} = \frac{\eta_{A1}F_{IM}(1 - \eta_{A2}\eta_{PR}) \pm \sqrt{\eta_{A1}\eta_{A2}F_{IM}L(1 + F_{IM})(1 - \eta_{A2}\eta_{PR})}}{\eta_{A1}F_{IM}(1 - \eta_{A2}\eta_{PR}) - \eta_{A2}L(1 + F_{IM})} \tag{B.63}$$

However, only the solution using the negative squareroot term leads to a practical coupling factor.

REFERENCES

[1] K.J. PARSONS AND P.B. KENINGTON. Effect of delay mismatch on a feedforward amplifier. *IEE Proceedings-G*, 141(2):140–144, April 1994.

[2] W.T.B. SLINGSBY. *An Antenna Isolation-Enhancing System for On-Frequency Radio Repeaters.* PhD thesis, Dept. of Electrical & Electronic Engineering, University of Bristol, October 1994.

[3] K.J. PARSONS AND P.B. KENINGTON. The efficiency of a feedforward amplifier with delay loss. *IEEE Transactions on Vehicular Technology*, 43(2):407–412, May 1994.

[4] P. B. KENINGTON. Efficiency of feedforward amplifiers. *IEE Proceedings-G*, 139(5):591–593, October 1992.

Appendix C

Amplifier and Predistorter Output P.D.F.s

The class C amplifier output p.d.f.s for the $(\pi/4)$–DQPSK and Rayleigh distributed input signals, derived in Section 5.5.2, are depicted in Figure C.1.

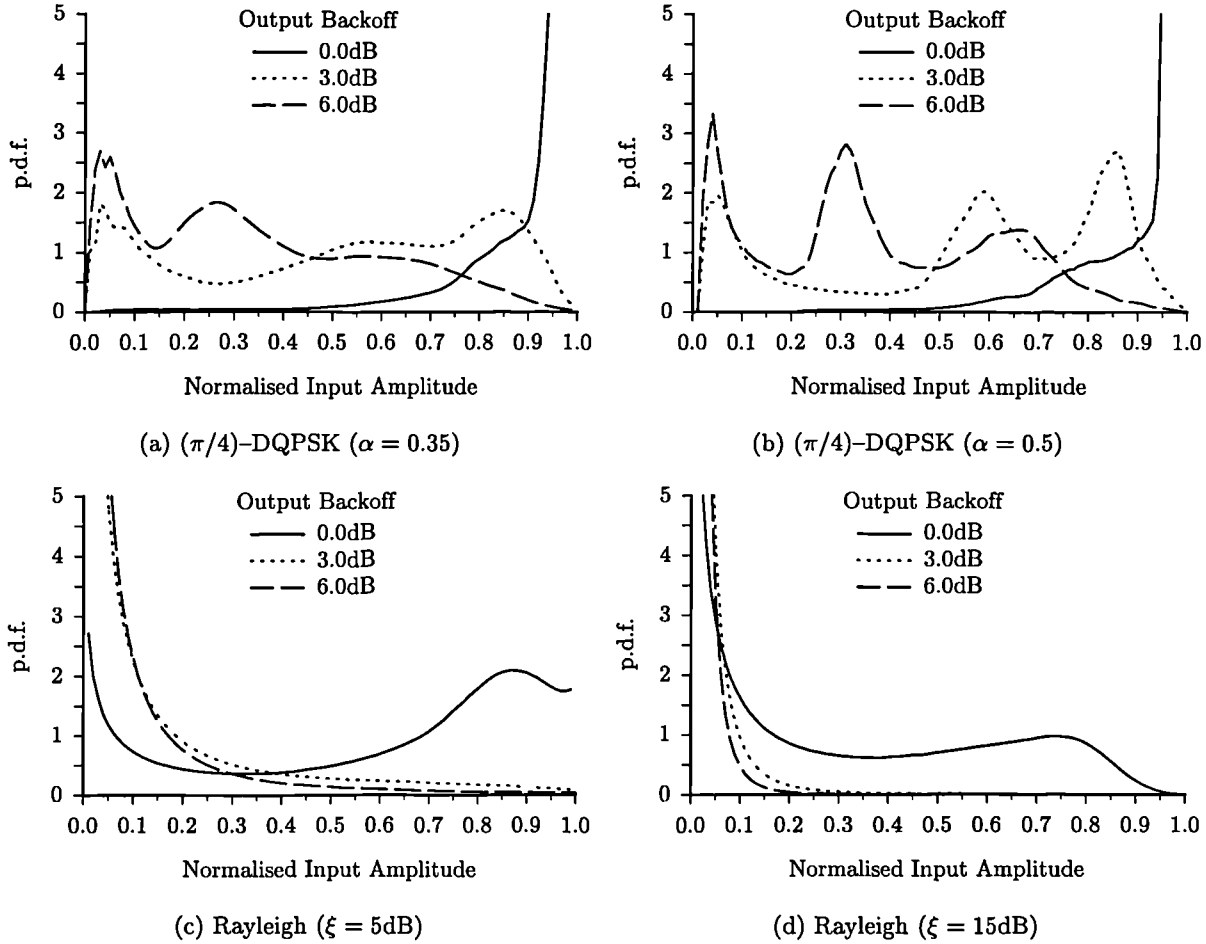


Figure C.1: Class C amplifier output p.d.f.s for various input signal p.d.f.s.

The ideal predistorter output p.d.f.s for the $(\pi/4)$–DQPSK and Rayleigh distributed input signals are depicted in Figure C.2.

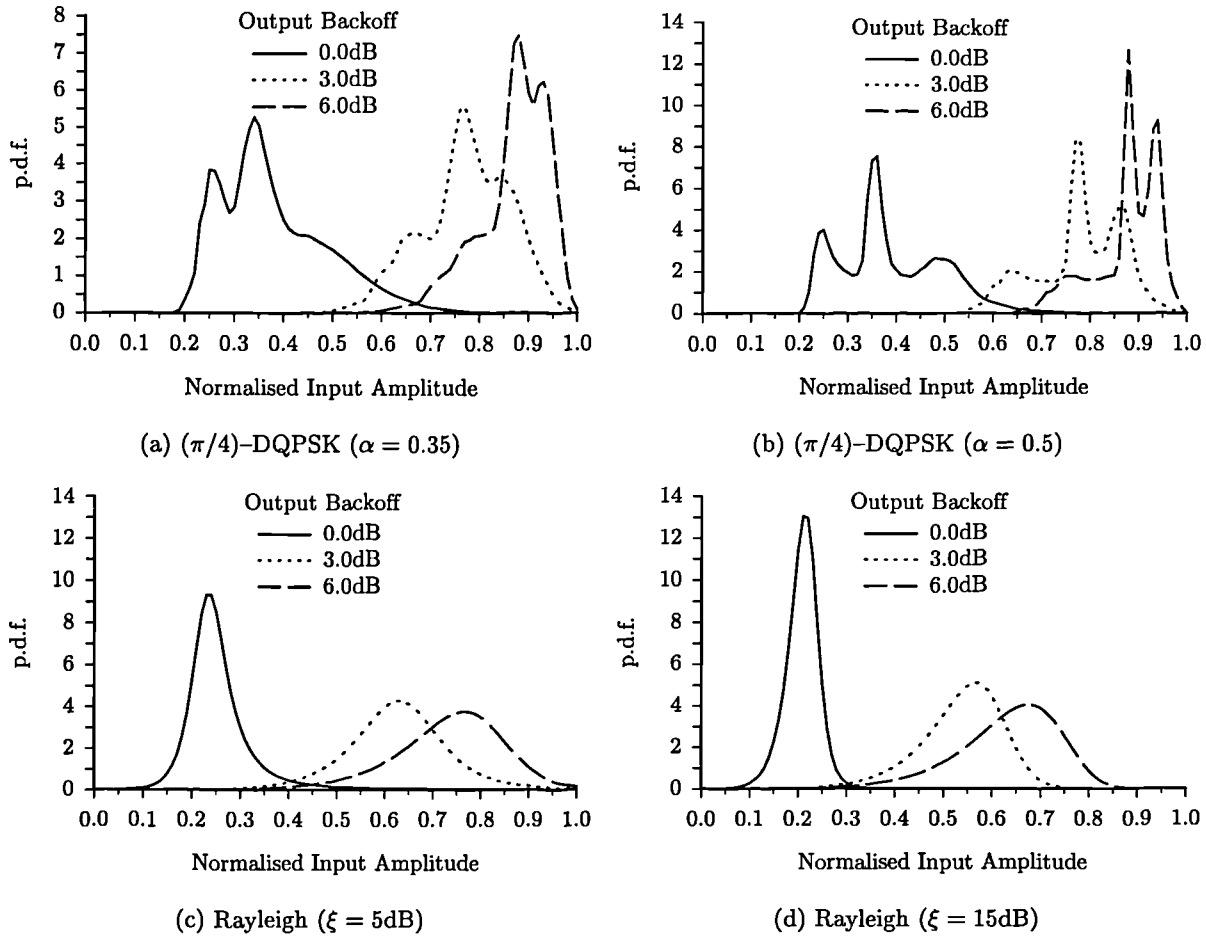


Figure C.2: Predistorter output p.d.f.s for various input signal p.d.f.s.

Appendix D

Operational Amplifier PLP Analysis

D.1 Monotonic–Increasing Circuit

The circuit for a monotonic–increasing operational amplifier PLP is shown in Figure D.1. The analysis assumes that all of the diodes switch on at a voltage of V_D, and in this state have zero resistance. At zero input signal, all of the diodes are assumed to be off. As the input voltage increases, each diode will switch on in turn. It is assumed that the first diode to turn on is D_1, then D_2, *etc.*, until the final diode, D_N, switches on.

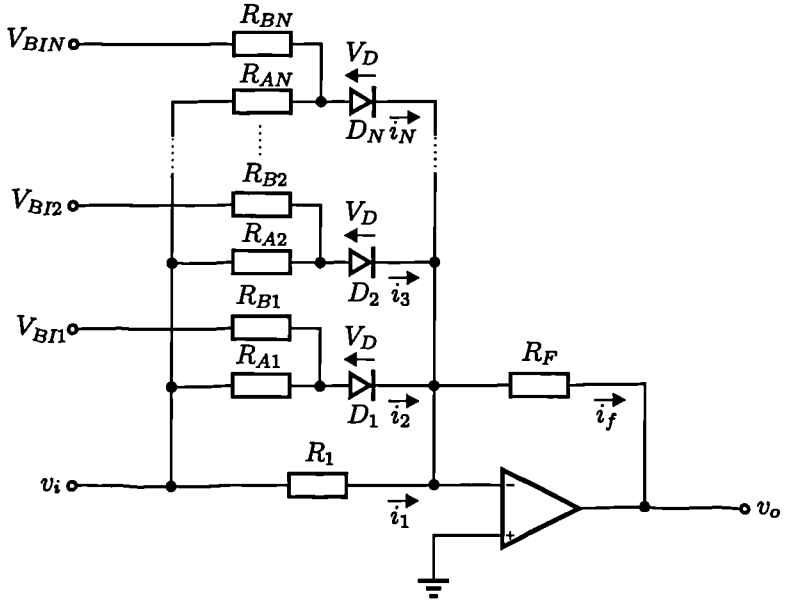


Figure D.1: Monotonic–Increasing Opamp based PLP.

When all of the diodes are off, the transfer characteristic is simply that of an inverting opamp

circuit, thus

$$v_o = \frac{-v_i R_F}{R_1} \tag{D.1}$$

The first diode, D_1, will switch on when the following condition is satisfied.

$$\frac{v_i - V_D}{R_{A1}} = \frac{V_D - V_{BI1}}{R_{B1}} \tag{D.2}$$

Where V_{BI1} is the bias voltage applied to the diode, via the bias resistor, R_{B1}, and is generally negative. Consequently, for the first diode to switch on, the input voltage must satisfy

$$v_i \geq \frac{V_D(R_{A1} + R_{B1}) - R_{A1}V_{BI1}}{R_{B1}} \tag{D.3}$$

The current flowing through the feedback resistor, i_f, is given by

$$i_f = i_1 + i_2 \tag{D.4}$$

Where

$$i_1 = \frac{v_i}{R_1} \tag{D.5}$$

$$i_2 = \frac{v_i - V_D}{R_{A1}} + \frac{V_{BI1} - V_D}{R_{B1}} \tag{D.6}$$

By substitution and simplification, the equation for the feedback current becomes

$$i_f = v_i\left(\frac{1}{R_1} + \frac{1}{R_{A1}}\right) + \frac{V_{BI1}}{R_{B1}} - V_D\left(\frac{1}{R_{A1}} + \frac{1}{R_{B1}}\right) \tag{D.7}$$

The output voltage of the PLP is given by

$$v_o = -i_f R_F \tag{D.8}$$

By substitution of Equation D.7, this yields

$$v_o = -R_F\left\{v_i\left(\frac{1}{R_1} + \frac{1}{R_{A1}}\right) + \frac{V_{BI1}}{R_{B1}} - V_D\left(\frac{1}{R_{A1}} + \frac{1}{R_{B1}}\right)\right\} \tag{D.9}$$

In a similar manner to the first diode, the second diode, D_2, will switch on when the input voltage satisfies

$$v_i \geq \frac{V_D(R_{A2} + R_{B2}) - R_{A2}V_{BI2}}{R_{B2}} \tag{D.10}$$

At this point diodes D_1 and D_2 are both on. Hence, the feedback current is now given by

$$i_f = i_1 + i_2 + i_3 \tag{D.11}$$

Where

$$i_3 = \frac{v_i - V_D}{R_{A2}} + \frac{V_{BI2} - V_D}{R_{B2}} \tag{D.12}$$

Hence, the transfer characteristic becomes

$$v_o = -R_F\left\{v_i\left(\frac{1}{R_1} + \frac{1}{R_{A1}} + \frac{1}{R_{A2}}\right) + \frac{V_{BI1}}{R_{B1}} + \frac{V_{BI2}}{R_{B2}} - V_D\left(\frac{1}{R_{A1}} + \frac{1}{R_{B1}} + \frac{1}{R_{A2}} + \frac{1}{R_{B2}}\right)\right\} \tag{D.13}$$

In general, the n^{th} diode will switch on when

$$v_i \geq \frac{V_D(R_{An} + R_{Bn}) - R_{An}V_{BIn}}{R_{Bn}} \tag{D.14}$$

The general expression for the transfer characteristic is given by

$$v_o = -R_F\left\{v_i\left(\frac{1}{R_1} + \frac{1}{R_{A1}} + \cdots + \frac{1}{R_{An}}\right) + \frac{V_{BI1}}{R_{B1}} + \cdots + \frac{V_{BIn}}{R_{Bn}} - V_D\left(\frac{1}{R_{A1}} + \frac{1}{R_{B1}} + \cdots + \frac{1}{R_{An}} + \frac{1}{R_{Bn}}\right)\right\} \tag{D.15}$$

Or

$$v_o = -R_F\left\{v_i\left(\frac{1}{R_1} + \sum_{i=1}^{n}\frac{1}{R_{Ai}}\right) + \sum_{i=1}^{n}\frac{V_{BIi}}{R_{Bi}} - V_D\sum_{i=1}^{n}\left(\frac{1}{R_{Ai}} + \frac{1}{R_{Bi}}\right)\right\} \tag{D.16}$$

D.2 Monotonic–Decreasing Circuit

The circuit for a monotonic–decreasing operational amplifier PLP is shown in Figure D.2. The analysis again assumes that all of the diodes switch on at a voltage of V_D, and in this state have zero resistance. At zero input signal, all of the diodes are assumed to be off. As the input voltage increases, each diode will switch on in turn. It is assumed that the first diode to turn on is D_1, then D_2, *etc.*, until the final diode, D_N, switches on

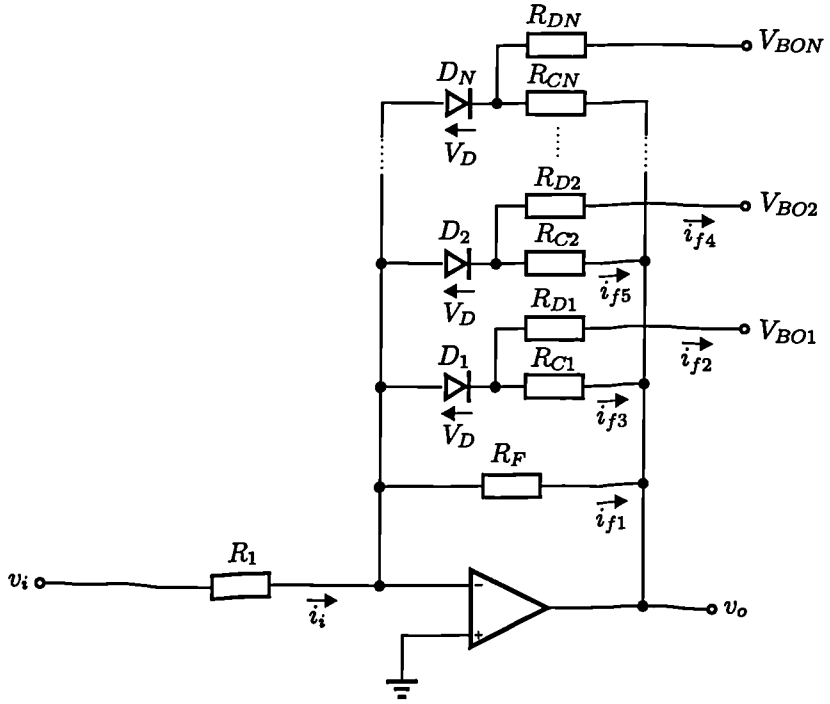


Figure D.2: Monotonic–Decreasing Opamp based PLP.

When all of the diodes are off, the transfer characteristic is simply that of an inverting opamp circuit, thus

$$v_o = \frac{-v_i R_F}{R_1} \tag{D.17}$$

The first diode, D_1, will switch on when the following condition is satisfied.

$$\frac{V_{BO1} + V_D}{R_{D1}} = \frac{-V_D - v_o}{R_{C1}} \tag{D.18}$$

Where V_{BO1} is the bias voltage applied to the diode, via the bias resistor, R_{D1}, and is generally positive. Consequently, for the first diode to switch on, the output voltage must satisfy

$$v_o \leq -R_{C1}\left\{\frac{V_{BO1}}{R_{D1}} + V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}}\right)\right\} \tag{D.19}$$

Thus, the knot abscissæ are controlled by the output voltage, rather than the input voltage, as is the case with the monotonic-increasing PLP circuit. The input voltage required to switch on the first diode must satisfy

$$v_i \geq \frac{R_1}{R_F} R_{C1}\left\{\frac{V_{BO1}}{R_{D1}} + V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}}\right)\right\} \tag{D.20}$$

The input current, i_i, is given by

$$i_i = i_{f1} + i_{f2} + i_{f3} \tag{D.21}$$

Where

$$i_{f1} = \frac{-v_o}{R_F}$$
$$i_{f2} = \frac{-V_D - v_o}{R_{C1}}$$
$$i_{f3} = \frac{-V_D - V_{BO1}}{R_{D1}}$$

Hence,

$$i_i = -v_o\left(\frac{1}{R_F} + \frac{1}{R_{C1}}\right) - \frac{V_{BO1}}{R_{D1}} - V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}}\right) \tag{D.26}$$

The input current is also given by

$$i_i = \frac{v_i}{R_1} \tag{D.27}$$

Equating these expressions and rearranging yields the transfer characteristic, thus

$$v_o = \frac{-1}{\left(\frac{1}{R_F} + \frac{1}{R_{C1}}\right)}\left\{\frac{v_i}{R_1} + \frac{V_{BO1}}{R_{D1}} + V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}}\right)\right\} \tag{D.28}$$

In a similar manner to the first diode, the second diode, D_2, will switch on when the output voltage satisfies

$$v_o \leq -R_{C2}\left\{\frac{V_{BO2}}{R_{D2}} + V_D\left(\frac{1}{R_{C2}} + \frac{1}{R_{D2}}\right)\right\} \tag{D.29}$$

This corresponds to an input voltage of

$$v_i \geq R_1\left\{R_{C2}\left(\frac{1}{R_F} + \frac{1}{R_{C1}}\right)\left(\frac{V_{BO2}}{R_{D2}} + V_D\left(\frac{1}{R_{C2}} + \frac{1}{R_{D2}}\right)\right) - \frac{V_{BO1}}{R_{D1}} - V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}}\right)\right\} \tag{D.30}$$

At this point diodes D_1 and D_2 are both on. Hence, the input current is given by

$$i_i = i_{f1} + i_{f2} + i_{f3} + i_{f4} + i_{f5} \tag{D.31}$$

Where,

$$i_{f4} = \frac{-V_D - v_o}{R_{C2}} \tag{D.32}$$

$$i_{f5} = \frac{-V_D - V_{BO2}}{R_{D2}} \tag{D.33}$$

Hence,

$$i_i = -v_o\left(\frac{1}{R_F} + \frac{1}{R_{C1}} + \frac{1}{R_{C2}}\right) - \frac{V_{BO1}}{R_{D1}} - \frac{V_{BO2}}{R_{D2}} - V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}} + \frac{1}{R_{C2}} + \frac{1}{R_{D2}}\right) \tag{D.34}$$

Consequently, the transfer characteristic is given by

$$v_o = \frac{-1}{\left(\frac{1}{R_F} + \frac{1}{R_{C1}} + \frac{1}{R_{C2}}\right)}\left\{\frac{v_i}{R_1} + \frac{V_{BO1}}{R_{D1}} + \frac{V_{BO2}}{R_{D2}} + V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}} + \frac{1}{R_{C2}} + \frac{1}{R_{D2}}\right)\right\} \tag{D.35}$$

In general, the n^{th} diode will switch on when the output voltage satisfies

$$v_o \leq -R_{Cn}\left\{\frac{V_{BOn}}{R_{Dn}} + V_D\left(\frac{1}{R_{Cn}} + \frac{1}{R_{Dn}}\right)\right\} \tag{D.36}$$

Corresponding to an input voltage of

$$v_i \geq R_1\left\{R_{Cn}\left(\frac{1}{R_F} + \sum_{i=1}^{n-1}\frac{1}{R_{Ci}}\right)\left(\frac{V_{BOn}}{R_{Dn}} + V_D\left(\frac{1}{R_{Cn}} + \frac{1}{R_{Dn}}\right)\right) - \sum_{i=1}^{n-1}\frac{V_{BOi}}{R_{Di}} - V_D\sum_{i=1}^{n-1}\left(\frac{1}{R_{Ci}} + \frac{1}{R_{Di}}\right)\right\} \tag{D.37}$$

In general, the transfer characteristic is given by

$$v_o = \frac{-1}{\left(\frac{1}{R_F} + \frac{1}{R_{C1}} + \cdots + \frac{1}{R_{Cn}}\right)}\left\{\frac{v_i}{R_1} + \frac{V_{BO1}}{R_{D1}} + \cdots + \frac{V_{BOn}}{R_{Dn}} + V_D\left(\frac{1}{R_{C1}} + \frac{1}{R_{D1}} + \cdots + \frac{1}{R_{Cn}} + \frac{1}{R_{Dn}}\right)\right\} \tag{D.38}$$

Or

$$v_o = \frac{-1}{\left(\frac{1}{R_F} + \sum_{i=1}^{n}\frac{1}{R_{Ci}}\right)}\left\{\frac{v_i}{R_1} + \sum_{i=1}^{n}\frac{V_{BOi}}{R_{Di}} + V_D\sum_{i=1}^{n}\left(\frac{1}{R_{Ci}} + \frac{1}{R_{Di}}\right)\right\} \tag{D.39}$$

D.3 Non–Monotonic Circuit

A non–monotonic operational amplifier PLP circuit can be created by merging the two previously described circuits. The transfer characteristic may be evaluated using the previous expressions for each PLP in turn.

To calculate the monotonic–increasing transfer characteristic expression, the feedback resistor value, R_F, is replaced by the effective feedback resistance, which is given by the parallel equivalent of R_F and all of the R_{Ci}'s which are currently switched in circuit.

To calculate the monotonic–decreasing transfer characteristic expression, the input resistor value, R_1, is replaced by the effective input resistance, which is given by the parallel equivalent of R_1 and all of the R_{Ai}'s which are currently switched in circuit.

ENGINEERING